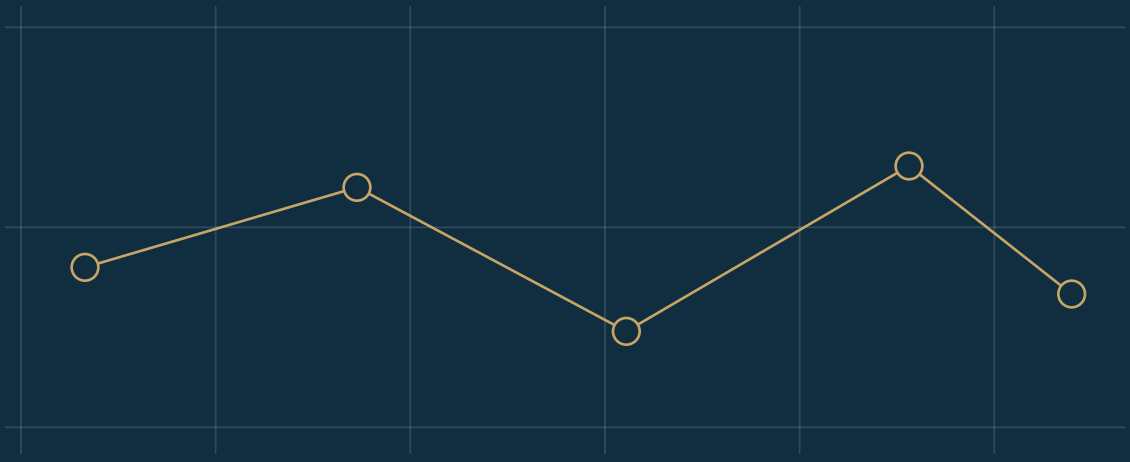
LUMINA LITERATI PUBLISHING

Artificial Intelligence, Law, and Governance in Pakistan

Muhammad Sohail Asghar
Hafsa Naz



OPEN ACCESS

Artificial Intelligence, Law, and Governance in Pakistan

Muhammad Sohail Asghar
Hafsa Naz

Lumina Literati Publishing
2025

Published 2025

Artificial Intelligence, Law, and Governance in Pakistan
Muhammad Sohail Asghar and Hafsa Naz

Published by Lumina Literati Publishing
luminaliterati.com

ISBN: 978-627-7813-63-5

DOI: 10.65758/isbn.9786277813635

Contents

Contents (continued)

List of Tables

About the Authors

Muhammad Sohail Asghar

Muhammad Sohail Asghar is an Assistant Professor of Law at the University of Okara, Pakistan, with more than 13 years of teaching and research experience. His academic work lies at the intersection of law, technology, and governance, with particular emphasis on artificial intelligence, digital privacy, constitutional rights, administrative law, and the legal challenges created by emerging technologies. His research examines how AI systems reshape traditional concepts of legal authority, responsibility, accountability, dignity, equality, and individual rights. In this book, he brings a predominantly legal and institutional perspective to the governance of artificial intelligence in Pakistan, examining constitutional competence, regulatory authority, public-sector deployment, judicial control, and rights-based safeguards.

Hafsa Naz

Hafsa Naz is an independent technology researcher whose work focuses on the development and responsible deployment of technology for the broader benefit of society. Her interests encompass emerging technologies, technological governance, digital systems, and the social implications of innovation. Her contribution to this book brings a technology-oriented perspective to questions of AI deployment, institutional capacity, data infrastructures, procurement, system reliability, and lifecycle oversight.

Together, the authors bridge legal and technological perspectives to explore practical pathways for responsible AI governance in Pakistan.

CHAPTER 01

Constitutional Competence and Institutional Authority over Artificial Intelligence

Introduction

Artificial intelligence (AI) has entered public law in Pakistan through the work of public administration. It has arrived through existing functions, powers, responsibilities, evidence and means of redress, without first becoming a distinct legal discipline or a legal entity of its own.

Before asking whether Pakistan needs an AI-specific statute or regulator, we need to settle some more fundamental questions. What do these systems do, and whose interests do they affect? Which level of government controls the underlying function? Which public authority may lawfully authorise, purchase, regulate or rely on the system? And which forum can intervene when its use produces error, abuse or exclusion?

This chapter establishes the constitutional and institutional basis for answering those questions.

The relevant framework remains underdeveloped, but it is misleading to describe it as a complete legal absence. The National Artificial Intelligence Policy 2025, formally adopted on 31 July 2025, sets national objectives and a programme for implementing them. The Digital Nation Pakistan Act 2025 also established the National Digital Commission and Pakistan Digital Authority. It authorises national master planning and coordination, with provision for AI oversight and standards where these are assigned under the Act. Neither development makes AI an independent constitutional source of authority or removes every existing sectoral mandate. We must still distinguish a policy commitment from statutory coordination, sectoral regulation, purchasing power and the authority to decide an individual case.

This distinction matters especially in a federal system. A common registry, a nationally purchased model or a centrally hosted platform may use the same technical architecture across healthcare, education, policing, labour relations and local services. Yet authority to create the relevant records and exercise discretion at the point of delivery may lie with different institutions. The residual allocation under the Eighteenth Amendment, the

constitutional relationship between legislative and executive competence, the Council of Common Interests and provincial responsibility for local government all bear on that allocation. Governance must therefore follow the public function. The geographical reach of the technology cannot settle the question by itself.

Section 1.1 defines AI by its functions, consequences and degree of meaningful human control, without tying the definition to a particular technology. Section 1.2 examines federal, provincial and local competence and develops a Federal Fit Test. Section 1.3 identifies the institutions that may authorise, supervise, investigate, audit, review or provide a remedy, and distinguishes their formal competence from demonstrated capability. Section 1.4 explains the different legal effects of constitutional provisions, statutes, delegated rules, precedent, policies, standards, licences, procurement agreements and administrative directions.

1.1 Artificial Intelligence as an Object of Legal Governance

Governing AI does not first require us to settle what intelligence means for every scientific or philosophical purpose. The legal task is more practical: identify the computational function, place it within an institutional process and examine what happens when its output affects people or public power. An excessively technical definition may become obsolete as architectures change. A definition that follows a vendor's label can be manipulated, while one covering every digital tool may hide the systems that need close scrutiny. The definition used here therefore follows function and consequence and remains neutral as to technology.

We should begin with the status of the present policy. The Government of Pakistan's Ministry of Information Technology and Telecommunication records the National Artificial Intelligence Policy 2025 as approved on 31 July 2025. It sets national goals and an implementation plan covering a public-sector register of AI activity, life-cycle and impact assessments, audits, algorithmic accountability, human oversight and redress. It remains a policy document. It does not enact AI-specific statutory provisions or supply generally applicable definitions of AI, and it gives no operational definitions of ‘critical’, ‘high risk’ or ‘high-impact’ AI. Its call for a future legal framework with corresponding punishments reinforces the distinction between present policy commitments and enacted duties (MoITT, 2025).

For this book, an AI-related system is an automated decision-making system, or a related system, that uses machine-based inference to produce predictions, classifications, rankings, recommendations, generated content, matches, anomalies, decisions or other outputs capable of affecting a physical or virtual environment. Close scrutiny is warranted where its operation may cause material harm through error, exclusion, surveillance, discrimination, security breaches or the absence of an effective remedy.

Relevant characteristics include the system's functions and scale, the population exposed, whether its effects can be reversed, the sensitivity of the information it processes, meaningful human participation in decisions and its relationship with governmental authority. These considerations are not exhaustive.

From product labels to legally relevant functions

'AI-powered', 'smart' and 'autonomous' make unreliable legal boundaries. Such labels describe products; they rarely explain the source or limits of public authority. A modest office aid and a system controlling access to public services may carry the same label. A rules engine or conventional statistical system may carry no AI label at all, yet rank, score or exclude people in ways that raise the same concerns about legality, equality, privacy, evidence and remedy as a learned model. Following the branding would burden some harmless tools while leaving consequential systems outside scrutiny.

The updated OECD definition offers a useful starting point for comparison. It describes a machine-based system pursuing explicit or implicit objectives and drawing inferences from inputs to generate predictions, content, recommendations or decisions that can influence physical or virtual environments. Autonomy and adaptability after deployment vary in degree; neither is simply present or absent (Organisation for Economic Co-operation and Development [OECD], 2024a). Article 3(1) of Regulation (EU) 2024/1689 of the European Parliament and of the Council, the Artificial Intelligence Act, adopts substantially the same structure (European Parliament & Council of the European Union, 2024, art. 3(1)). This convergence helps us classify systems for comparative purposes. Neither instrument determines Pakistani constitutional authority. The OECD formulation provides a broad technical starting point, not a complete account of legal scope. Its output categories must still be connected to the functions of the institutions using them.

Several AI functions may operate in sequence. A system may classify an application, convert that classification into a risk score, place the application in a queue, recommend action and generate reasons for a disposition that initiates an administrative step. We must examine the validity of the data at every stage. A later calculation may run correctly yet still produce a faulty ranking because the initial classification was wrong. A recommendation may accurately reflect the model's assessment of an applicant without establishing that the official had authority to rely on it or that the assessment used the applicable legal criteria. Generated reasons may correctly describe the stated basis for a disposition while concealing the absence of the official's independent judgment. Each situation calls for an inquiry into lawful authority, adequate evidence, freedom from undue pressure or influence, and fairness of consequences. The OECD also identifies relevant functions: recognition and categorisation; identification and detection of events

or anomalies; forecasting and personalisation; support for interaction; optimisation; and knowledge-based reasoning (OECD, 2024a).

The book also covers related automated decision systems where a disputed technical boundary would defeat the legal purpose of classification. A deterministic rules engine may fall outside some definitions of AI and still cause the same constitutional or administrative injury as an inferential model. Including it does not decide whether it is lawful or impose identical requirements on every component. It allows materially equivalent exercises of automated power to receive comparable scrutiny, despite differences in their underlying software.

Degrees of influence and meaningful human control

These systems range from low-consequence assistance through decision support, recommendations, prioritisation and risk scoring to fully or effectively automated action. The categories describe what an output does in practice. They do not measure how sophisticated the technology is.

Low-consequence assistance helps a user complete a task without ordinarily settling a material issue. Decision support organises or analyses information, leaving the user to frame and decide the question. A recommendation proposes an option. Prioritisation orders people or cases and thereby distributes attention, time or resources. Risk scoring combines selected variables into an assessment that may influence eligibility, scrutiny, intervention or treatment. Fully automated action takes a consequential step without a fresh human determination of the particular case. Action is effectively automated where a person formally approves the output but has no genuine capacity to evaluate it and reach a different result.

A tool's position on this spectrum depends on its use. A generated summary may offer assistance when checked against the record, support a decision when it organises relevant information, or determine the outcome when its conclusions and reasons are routinely accepted. Workload, interface design, performance incentives and organisational practice may turn a recommendation into the usual decision by making departures exceptional. A procurement description or workflow diagram helps explain the intended role, but it cannot establish the classification. We need to examine the actual decision-making environment.

Comparative frameworks illustrate the point without determining Pakistani law. The voluntary Artificial Intelligence Risk Management Framework (AI RMF 1.0) of the U.S. National Institute of Standards and Technology (NIST) places human-AI interaction on a continuum from entirely manual to completely automated. It calls for roles, responsibilities, operator competence and oversight procedures to be defined at each level (Tabassi, 2023). Article 14 of the European Union's Artificial Intelligence Act takes a similar approach to oversight of high-risk systems. An effective overseer should

understand the system's capabilities and limits, detect anomalies, recognise automation bias and interpret its output. The overseer should also be able to disregard or modify an output, reverse changes where applicable and stop operation safely (European Parliament & Council of the European Union, 2024, art. 14). These functions are useful for analysis, although the EU's categories and duties do not apply to Pakistan by their own force. Human involvement must mean more than clicking through screens and signing at the end. The responsible person needs intelligible information about the output and its limits, relevant knowledge and experience, enough time for review, and direct access to all records and documentation relating to the process. That person must have lawful authority to depart from a recommendation and the practical ability to override the system, supported by procedures for recording reasons and reporting uncertainty. The institution must permit disagreement. Research on automation bias also suggests that an apparently authoritative output can shape an otherwise independent decision-maker's thinking even when that person remains 'in the loop' (Pfister, 2022). Formal authority may therefore exist without meaningful control.

This is a demanding test where an output directly affects public authority. An official cannot provide genuine oversight when decisive variables are hidden, the model's limitations are unavailable, changing the result is prohibited or no review record is retained. The official need not reproduce every computation. What matters is a sufficient basis, and the practical capacity, to decide whether reliance is lawful and appropriate in the individual case. Where these conditions are missing, a score or recommendation should be treated as effectively determinative even if a person performs the final formal act.

The operational chain and attributable responsibility

An AI-enabled outcome rarely comes from a model alone. A database supplies or stores information. A model processes selected inputs, and an interface presents, orders or withholds parts of the output. A professional or administrative user interprets it before an institution issues, implements or records the resulting act. Procurement, configuration, validation, maintenance, updates, cybersecurity and vendor support shape this sequence. We should scrutinise these elements as one operational chain, while recognising that they do not become a new legal person called 'the AI'.

Responsibility must remain traceable throughout a project. We need to distinguish the organisation issuing a programme from the agency purchasing its technical components and from the supplier or operator providing them. Responsibility for data provenance and quality must also be distinguished from responsibility for selecting and testing models, configuring interfaces, designing output formats, using the system professionally and taking the final action. Several parties may share responsibilities, provided the allocation is clear. Saying 'the system made that decision' tells us neither

who held lawful authority nor who accepted the output or whose conduct can be reviewed.

Attribution requires records. A person challenging a decision will likely need to know what data, model version, parameters and settings were used; how results were displayed; whether the user could depart from them; what review occurred; and what action followed. Aggregate performance claims, general descriptions of model operation and vendor assurances cannot reconstruct that particular chain. The book's reviewability test requires documents connecting authorisation, purchase, deployment, later changes, presentation and human use of outputs, final action, complaint and correction. The National AI Policy 2025 supports this approach through its proposals for a public-sector registry, life-cycle evaluations, impact assessments for high-risk systems, audits, algorithmic accountability and redress (MoITT, 2025). That is substantial policy support. The precise legal effects still depend on lawful implementation by competent governmental authorities.

Outsourcing leaves the question of public responsibility to be answered. A vendor may design, host, update or monitor a system, but a contract cannot give the purchasing institution a public power it does not possess. Technical opacity should not prevent that institution from explaining its decisions, preserving records, supervising changes, investigating failures, enabling review or implementing remedies. A professional must also make the assessment required by the applicable role before treating a technical output as a lawful professional judgment. Breaking the process into technical components should help us allocate responsibility. It should not make responsibility disappear.

Policy, standards, and enforceable obligation

Governance instruments have different legal effects, which cannot be settled by treating them as interchangeable rules or arranging them solely by title. The Constitution of the Islamic Republic of Pakistan, statutes, valid delegated legislation and binding precedent supply legal norms within their respective fields. Rules of business allocate executive work and responsibility. Administrative directions may bind officials or guide implementation where a competent legal source allows this. Strategies, white papers and roadmaps set objectives; codes and standards supply practices or benchmarks. Licences impose conditions under an enabling authority. Procurement contracts, service-level terms, data-sharing clauses and specifications bind through the legal powers and relationships under which they are made. For each instrument, we must identify its source, addressee, enabling authority, mandatory wording, incorporation method, enforcement route and relationship with superior law.

The National Artificial Intelligence (AI) Policy 2025 shows why these distinctions matter. Its official approval has governmental and administrative significance. Its

technical content can guide planning, implementation priorities, procurement and capacity building. Approval does not, however, make it a statute, create offences or enlarge a regulator's jurisdiction. Nor does it remove an available appeal or review. The policy identifies ‘critical’, ‘high risk’ and ‘high impact’ systems as possible subjects of greater regulatory control, but those categories remain undefined. They cannot, on their own, establish a legally defined threshold for regulation.

Technical standards require the same care. OECD, NIST and EU materials can help define terms, draft procurement requirements, assess regulatory impacts and review compliance. A standard may acquire binding force through a statute, delegated rules, an authorised licence condition or a contract. It may also provide a useful benchmark without becoming law. Technical authority alone gives it no domestic legal effect. International comparison should improve the questions we ask of a Pakistani system while leaving the answer to Pakistan's Constitution, legislation, precedent, valid administrative action and applicable contracts. An external institution's authority cannot replace those sources.

Dates and documentary status also matter. Since approval of the national policy, older claims of a general absence of national AI guidance are no longer accurate. Umer et al. (2023) remain relevant as a dated call for clearer healthcare guidance. Their review cannot establish the present position of every regulator or sector. An announcement does not prove procurement, procurement does not prove a pilot, and a pilot does not prove continuing operation. Even operation establishes neither lawful authorisation nor effective oversight. This chapter therefore distinguishes announced, proposed, procured, piloted, operational, suspended and unverified systems. Where a public instrument or activity cannot be identified, we record documentary uncertainty. A finding of absence requires an appropriately comprehensive inquiry.

A high-consequence scope test

The book uses a two-stage rule to determine scope. First, what does the system do, and where does its output enter the operational chain? Second, does this use require closer scrutiny? The latter inquiry considers the interests affected, the probability and severity of error, the number and vulnerability of people exposed, informational sensitivity, the persistence and reversibility of effects, cumulative or systemic scale, meaningful human control and the connection with public authority. NIST likewise combines the probability of harm with the magnitude of its consequences. The MAP function of AI RMF 1.0 calls for records of the task, deployment setting, likely impacts, limits of knowledge, intended use of outputs and human oversight (Tabassi, 2023). This supplies a comparative way to examine consequences, not a rule of Pakistani jurisdiction.

Six categories help apply the test consistently. Function asks whether the system predicts, classifies, ranks, recommends, generates, matches, detects anomalies, scores risk

or takes action. It also examines how closely the output is connected to the consequential outcome that follows.

Scale examines the range of cases, the geographical area affected and the time taken by the process. It also asks whether a problem arising in one area could spread into others.

Affected Population considers how many people may be exposed and whether they can limit that exposure. We must also ask what they know about the system and whether they can dispute or question the action taken.

Reversibility asks whether an error can be identified and corrected quickly. An error may instead cause lasting exclusion, loss of status or reputation, formal investigation, or repeated disadvantage to the same person.

Informational Sensitivity examines the information collected or inferred during use. Could its misuse or disclosure cause harm? Could information generated by the system come to be accepted as established fact?

Connection with Public Authority asks whether a system performs a public duty or shapes its discharge, distributes a public benefit or liability, directs scrutiny, authorises enforcement or alters the relationship between an institution and the people subject to its authority.

Meaningful human involvement affects every category throughout the process. Its role cannot be reduced to adding an appearance of credibility at the end.

'High-consequence' is an analytical term used in this book. It is not equated with the policy's undefined 'high-risk' or 'high-impact' expressions, or with the EU Artificial Intelligence Act's statutory risk categories. Classification does not decide that a system must be prohibited. It prompts closer examination of lawful authority, responsible actors, data and model records, validation, monitoring, human control, security, independent scrutiny, access to complaints and effective remedies. Examining function and consequence in this way helps identify the actual difficulty: missing authority, uncertain responsibility, inappropriate delegation, inadequate access to evidence, poor coordination or weak capacity. That diagnosis should come before proposals for a new institution or statute.

Section 1.2 turns from what is governed to who may govern it. Once we identify a system's function, consequences and degree of human control, we must locate the authority responsible for the underlying field. Which federal, provincial, sectoral or local institution may regulate it, authorise the public function, procure or operate the system, and review the resulting act?

1.2 Constitutional Allocation of Public Authority in a Federal Order

Artificial intelligence does not define its own field of legislative competence in Pakistan. The Constitution supplies no abstract legislative head for 'AI', 'digital technology', 'data' or 'cybersecurity'. A diagnostic model, admissions-ranking system, land-record anomaly detector and policing tool may use similar computational methods while serving very different public functions. We must locate authority through the function, the institution acting, the territory affected and the Fourth Schedule. Product labels and the reach of a network cannot perform that constitutional task.

Technical centralisation can make legally divided authority difficult to see. Federal funding, national hosting or a single interface may support outputs used in provincial health, education, policing, labour, land or municipal decisions. These arrangements do not transfer the underlying function. Provincial responsibility also leaves room for defined federal roles in communications, defence, federal institutions, specified economic regulation, national planning and inter-provincial coordination. Each component requires its own inquiry: who may legislate and administer, who holds and corrects the record, who may act on the output, and who must answer a challenge?

1.2.1 Legislative and executive competence

Articles 141–143 establish the basic legislative structure. Article 141 permits Parliament to legislate for all or part of Pakistan, including extra-territorially, and a Provincial Assembly to legislate for its Province. Territorial reach does not itself confer authority over a subject. Article 142(a) gives Parliament exclusive competence over the Federal Legislative List (FLL). Under Article 142(b), Parliament and Provincial Assemblies may both legislate on criminal law, criminal procedure and evidence. Subject to those concurrent fields, Article 142(c) reserves non-FLL matters to the Provinces. Article 142(d) gives Parliament exclusive competence for parts of Pakistan outside a Province. Under Article 143, a competent federal law prevails over an inconsistent provincial law; this does not validate federal legislation lacking a constitutional head of power (Constitution of the Islamic Republic of Pakistan, 1973, arts. 141–143 & Fourth Schedule).

Executive authority follows a related distribution. Article 97 extends federal executive authority to subjects on which Parliament may legislate. Unless the Constitution or a valid federal law expressly provides otherwise, that authority does not extend to a subject on which a Provincial Assembly may legislate. Article 137 similarly links provincial executive authority to provincial legislative competence, subject, in concurrent fields, to federal executive authority expressly conferred by the Constitution or a valid federal statute (Pakistan's Constitution of the Islamic Republic of Pakistan, 1973. Arts. 97,

137). A federal ministry therefore cannot assume power to decide an individual's provincial case merely because it writes policy, hosts a server, owns a model or pays a vendor. Every consequential legal act needs a competence map. Data collection and exchange, technical procurement, sector rules, model operation, reliance on outputs and decisions in individual cases may each involve a different authority. A national agency may lawfully host infrastructure while a provincial department retains sole decision-making power. A federal regulator may supervise the communications service used by a hospital or school without regulating the hospital or school itself. Describing all this as 'the government' deploying AI leaves the constitutional analysis unfinished.

Constitutional authority must also be distinguished from administrative capacity. Drawing on sixteen interviews, Ahmed et al. (2025) describe fragmented and under-resourced civilian cyber governance, including dependence on military-linked expertise and donor-driven technologies. Such findings may help explain who leads a programme. They do not change Articles 97, 137 or 142.

1.2.2 Defined federal heads and their limits

Several entries in FLL Part I may support defined AI-related measures. Entry 1 covers defence, the armed forces, the Federal Intelligence Bureau and preventive detention for reasons of State connected with defence, external affairs or Pakistan's security. Entries 3 and 32 concern external affairs and treaties; entry 4 covers nationality and citizenship; and entry 7 covers posts, telephones, wireless, broadcasting and similar communications. Entries 11, 13, 14 and 16 concern federal public services, federal institutions and related administrative matters. Intellectual property falls under entry 25. Entry 27 covers imports and exports, inter-provincial and foreign trade, and export-quality standards, while entries 28–31 address banking, insurance, cross-provincial markets and specified corporations. Entries 34–35 concern strategic roads and federal surveys; entry 39 addresses weights and measures; and entry 41 covers elections to the Presidency and federal and provincial legislatures, with related election institutions. Entry 57 permits inquiries and statistics for Part I purposes. Entry 58 concerns matters within Parliament's competence or relating to the Federation, and entry 59 covers matters incidental or ancillary to the enumerated Part I subjects (Constitution of the Islamic Republic of Pakistan, 1973, Fourth Schedule, FLL pt. I, entries 1, 3–4, 7, 11, 13–14, 16, 25, 27–32, 34–35, 39, 41, 57–59).

These provisions may support regulation of an algorithmic function in delivering a service, security controls for a federal system, analytical support for a federal institution or supervision of banking automation. The relevant entry must still cover the particular subject. None supplies a general right to access or manipulate every internet-connected database. Nor may an entry be used to absorb a field that would otherwise remain provincial. Its application must be justified by the function being regulated.

FLL Part II provides bases of authority with a distinct institutional arrangement. These cover federally controlled institutions and declared industries (entry 3), federal regulatory bodies (entry 6), and national planning and economic coordination, including coordination of science and technology research and development (entry 7). They also cover the Census (entry 9), specified police extensions across provincial boundaries and railway areas (entry 10), legal, medical and other professions (entry 11), standards for higher education and scientific institutions (entry 12), inter-provincial coordination (entry 13), and the Council of Common Interests (CCI) (entry 14). Under Article 154, the CCI must formulate and regulate policy for Part II and supervise and control the related institutions. The provision does not simply assign these tasks to another federal division (Constitution of the Islamic Republic of Pakistan, 1973, articles 153–154 and Fourth Schedule, FLL pt. II, entries 3, 6–7, 9–14).

Article 99(3) permits rules for allocating and conducting federal business. In the Rules of Business, 1973, as amended up to 24 March 2025, rule 3 and Schedule II, item 17 allocate IT planning, telecommunications coordination, government software standardisation and NITB-related responsibilities to the Information Technology and Telecommunication Division (Government of Pakistan, Cabinet Division, 2025). These provisions identify the federal portfolio that may act where the Federation is competent. They do not establish that competence in the first place.

Federal procurement depends on an existing mandate in the same way. The Public Procurement Regulatory Authority Ordinance, 2002 defines federal 'procuring agencies' and procurement within the Authority's scope (ss. 2(j)–(l)); the Public Procurement Rules, 2004 regulate procurement in that field. They govern how a competent federal institution acquires an AI system. They do not authorise its underlying public function, and they do not govern provincial procurement, which follows the relevant provincial regime. A lawful mandate supports procurement; procurement supplies no independent constitutional head of power.

1.2.3 Provincial fields and constitutional boundary points

Article 142(c) reserves subjects outside the FLL to Provincial Assemblies. Health, school education, ordinary police organisation, labour administration, land and municipal services therefore remain provincial to the extent that a particular measure falls outside the FLL and Article 142(b)'s concurrent fields. Work on mental-health law and post-devolution education describes the implementation of the Eighteenth Amendment's devolution (Ahmad et al., 2025; Tareen & Tareen, 2016). That literature provides context. The operative allocation rests on the Constitution and the relevant provincial legislation.

Using a national technical system in these fields does not move the substantive decision-making power. A centrally hosted model may prioritise patients, flag pupils,

recommend police deployments, select labour inspections or detect inconsistencies in land records. The provincial institution must still establish its authority to adopt the tool, determine the legal effect of the output and decide the case. The federal institution must separately identify the head supporting its contribution. Interoperability may be useful nationally, but a shared technical layer cannot itself give the Federation exclusive control of the service that uses it.

Other constitutional provisions limit broad jurisdictional claims. Article 151 addresses the free movement of goods across Provinces and the conditions for restrictions. It does not give the Federation general jurisdiction over data merely because digital services cross borders. Article 159 concerns broadcasting and unreasonable federal refusal to entrust certain broadcasting responsibilities to a Province. Even within an enumerated communications field, provincial participation may therefore receive constitutional protection. Article 164 allows the Federation or a Province to make grants for purposes outside its legislature's competence. Funding and lawmaking are distinct powers. A federal grant for a provincial digital-health programme would not make health a federal field (Constitution of the Islamic Republic of Pakistan, 1973, Articles 151, 159, 164).

Claims based on security require equal care. Article 260's definition of 'security of Pakistan' expressly excludes 'public safety as such'. A cybersecurity measure may still be justified within a competent federal field. Merely attaching a security label to a local database, however, cannot turn ordinary policing or service administration into a federal security function (Constitution of the Islamic Republic of Pakistan, 1973, Article 260). We must preserve this distinction when examining dual-use technology. A common platform may support a legitimate federal-security function alongside ordinary civilian work, but the shared platform does not erase the difference between those functions.

Capacity and finance matter without creating jurisdiction. Earlier studies of National Finance Commission Awards and vertical fiscal imbalance document longstanding asymmetry in public finances (Ara & Sabir, 2010; Ahmed et al., 2007). They establish neither later fiscal conditions nor the capacity to oversee a particular AI system. Their value here is historical: legal accountability and practical capability may diverge. Capacity support should enable a Province to oversee its system. Financial dependence should not leave it answerable for outputs that it cannot monitor, modify or suspend.

1.2.4 Local government and vertical digital administration

Article 140A requires every Province to establish local government by law and devolve political, administrative and financial responsibility and authority to elected local representatives. It mandates devolution without prescribing one national list of local functions. Whether a local body controls a health facility, school process, municipal record, land-related service or complaint route depends on the provincial local-government and sector statutes in force at the relevant time (Constitution of the

Islamic Republic of Pakistan, 1973, art. 140A). A national AI programme should therefore avoid treating 'local government' as a functionally identical tier throughout Pakistan.

Digital administration across levels of government nevertheless raises recurring problems. Local staff may create or verify a source record, a Province may set eligibility rules, and a national platform may exchange information. A vendor may run the model while a frontline official communicates the result. Responsibility must be attributed at every link. Hosting a registry does not authorise decisions about entitlement. Prescribing a data format does not permit removal of local discretion granted by provincial law. An officer's electronic confirmation provides no meaningful control where the interface hides reasons, prevents correction or makes departure from the output practically impossible.

Research on local education administration and the exercise of discretion shows that implementation depends on frontline interpretation, constraints and adaptation as well as policy text (Farooqi & Forbes, 2020; Komatsu, 2009). Centralised systems may improve consistency and portability. They may also fix categories that fit some districts poorly, treat missing data as evidence of ineligibility, or leave the office receiving a complaint unable to correct the record. Federal fit must therefore address record provenance, correction authority, override procedures, local language and infrastructure. The complaint route must reach the institution that controls the error.

Article 150 requires full faith and credit throughout Pakistan for each Province's public acts and records. It does not authorise their collection, exchange, linkage or automated use.

1.2.5 Coordination under the Constitution and the Digital Nation Pakistan Act, 2025

Articles 144–149 provide defined routes for intergovernmental action, each with limits. Under Article 144, one or more Provincial Assemblies may resolve that Parliament should legislate for them on a non-FLL subject. Each consenting Province may later amend or repeal that Act as it applies there. Article 145 concerns a Governor acting as the President's agent for federal areas outside a Province; it gives no general power to direct provincial administration. Article 146 permits consensual entrustment of federal functions to a Province. It also allows a competent federal Act to confer powers or duties on provincial authorities, with compensation for additional administrative costs. Article 147 permits provincial functions to be entrusted to the Federation, subject to Provincial Assembly ratification within sixty days (Constitution of the Islamic Republic of Pakistan, 1973, arts. 144–147).

Article 148 requires provincial executive authority to secure compliance with applicable federal law. When the Federation exercises executive authority in a Province, it

must have regard to that Province's interests. Article 149 permits directions in specified circumstances: preventing prejudice to federal executive power; requiring construction and maintenance of communications declared nationally or strategically important in the direction; and addressing a grave menace to Pakistan's peace, tranquillity or economic life. It supplies no general digital-coordination power (Constitution of the Islamic Republic of Pakistan, 1973, arts. 148–149). An arrangement must identify its lawful route, whether divided administration, consent, entrustment, competent federal legislation or a narrowly authorised direction.

The CCI is the constitutional coordination forum for FLL Part II. Chaired by the Prime Minister, it includes every Chief Minister and three federal nominees. It is responsible to Parliament and must submit an annual report. It has a permanent secretariat, must meet at least once every ninety days and decides by majority. A dissatisfied government may refer a decision to Parliament in joint sitting (Constitution of the Islamic Republic of Pakistan, 1973, arts. 153–154). Research on Centre–Sindh relations in 2018–2022 illustrates how executive federalism works as a practical relationship as well as a formal arrangement. One Province over one period cannot establish an enduring national pattern (Faiz, 2023).

The Digital Nation Pakistan Act, 2025 establishes machinery for planning, coordination, standards and compliance. Sections 3–5 create the National Digital Commission (NDC), whose membership includes the Chief Ministers. Section 8(b) empowers PDA to issue and enforce instruments implementing the Masterplan. Section 8(c) requires PDA to facilitate coordination among federal, provincial, local, sectoral and regulatory institutions and private stakeholders. Under section 8(g), PDA may undertake AI oversight, standards and compliance work only in areas assigned by the NDC, without overlapping other regulators or conflicting with existing law. Section 8(h) preserves designated data custodians' control and responsibility, while section 11(4) requires federal-provincial consultation. Under section 12, identified entities prepare alignment plans. PDA monitors and reports compliance, may recommend correction or sanctions, and must formally notify a non-compliant public entity before escalating the matter to the Commission (Digital Nation Pakistan Act, 2025, ss. 3–5, 8(b)–(c), 8(g)–(h), 11(4), 12).

These mechanisms have real statutory regulatory and compliance force. They do not give the Federation universal jurisdiction over technology. Section 28 gives the Act priority where inconsistency arises, subject to specified data-rights and cybersecurity laws and its consultation process. Constitutional supremacy still prevents the Act from creating a missing FLL head. Displacing or entrusting a provincial sector function requires a separate constitutional route and substantive legal basis, while the CCI retains its constitutional role in Part II matters. Section 8(h) preserves each designated custodian's control and responsibility for its dataset. Responsibility for a decision or

digital service must be traced separately through section 8, the Masterplan and applicable sector law. NDC and PDA processes can assist that allocation; they cannot replace it.

The Twenty-seventh Amendment changed constitutional forums and FLL Part I entry 55 without reallocating the substantive sectoral heads considered here. Article 175E(1) gives the Federal Constitutional Court exclusive original jurisdiction over disputes between two or more federal or provincial governments. Article 175E(2) confines judgments in those disputes to declarations. Article 175F provides constitutional appeals, including appeals by leave from Article 199 decisions. Where Article 202A is in force, its Constitutional Benches exercise Article 199 jurisdiction. Amendments to Articles 146(3), 152 and 159(4) assign specified appointments of arbitrators to the Chief Justice of the Federal Constitutional Court (Constitution of the Islamic Republic of Pakistan, 1973, arts. 146(3), 152, 159(4), 175E–F, 199, 202A; Constitution (Twenty-seventh Amendment) Act, 2025, ss. 15–17, 21). Governance arrangements should settle routine disputes administratively while recognising these judicial routes.

1.2.6 A federal-fit test for AI arrangements

Before procurement or deployment, a high-consequence system involving more than one level of government should satisfy eight connected inquiries.

First, establish competence for each consequential act. The proponent should identify the relevant FLL entry or residual provincial subject, the competent legislature, executive authority under Article 97 or 137, and the sector instrument supporting data collection, procurement, operation and decision. Where cooperation relies on Article 144, 146 or 147, the necessary resolution, consent, entrustment, ratification and cost arrangements should be recorded. A strategy, Masterplan, grant or nationwide presence cannot supply missing legal authority.

Second, distinguish the public function from its technical components. Hosting, cybersecurity, identity verification, model operation, standard-setting and substantive adjudication may fall to different institutions. The arrangement should identify the power attached to each component and prevent a technical operator from gaining control over a sector decision without scrutiny. A federal platform may support provincial administration while the Province remains the legal decision-maker.

Third, connect responsibility with practical control. A Province or locality should not bear sole responsibility if it cannot inspect the relevant documentation, correct records, alter thresholds, demand remediation or suspend use. Equally, a central institution should not disclaim responsibility where mandatory classifications or platform controls effectively settle outcomes. The agreement should identify who approves changes, monitors performance, records departures and answers for each type of failure.

Fourth, govern information flows and custodianship. The design should identify the authoritative record, designated custodian, lawful recipients, correction process, retention rules and institution responsible for explaining an output. Interoperability gives no general permission to transfer data. Central coordination should preserve each designated custodian's control and responsibility for its dataset, consistently with section 8(h) of the Digital Nation Pakistan Act, 2025. Responsibility for decisions and services requires a separate allocation.

Fifth, give intergovernmental participation legal substance. Provinces should participate in defining the problem, selecting safeguards, validating performance measures and approving material changes while the design is still open. Part II policy and related institutional supervision require the CCI's constitutional role. Non-FLL legislation or entrusted functions require the appropriate route under Article 144, 146 or 147. Participation in the NDC and Masterplan consultation provide an additional statutory forum; they cannot substitute for those constitutional requirements.

Sixth, preserve lawful local adaptation. The applicable provincial statute should determine local functions and discretion. System design should specify when frontline staff may depart from an output, which reasons they may consider, how local conditions enter the assessment and how residents can correct an error. Uniformity fails the federal-fit test if it quietly removes authority devolved by provincial law under Article 140A.

Seventh, connect finance, procurement and capacity to the proper mandate. The responsible level of government needs continuing funds, trained staff, audit access, security capability, control over changes and an exit route, as well as the initial software. Federal PPRA rules govern federal agencies; provincial procurement follows the relevant provincial law. Article 164 permits funding outside the funder's legislative field, but the grant should strengthen the competent institution's ability to act without displacing it.

Eighth, specify dispute resolution, review and remedy. Disagreements between governments about competence, costs, standards, data or changes need an agreed administrative route consistent with CCI and Federal Constitutional Court jurisdiction. Individuals need a direct way to correct records and contest decisions. A complaint must not circulate among the local office using an output, the Province providing the service, the federal platform operator and a vendor invoking confidentiality. The arrangement must name the institution that receives, investigates, decides and remedies each kind of grievance.

For a national health-risk system, the test would ask whether the Federation is coordinating standards or seeking to control provincial clinical administration. Did the Provinces approve the data and performance rules? Can facilities correct records and depart from recommendations, and who answers the patient? For education prioritisation, we would ask whether a national objective has become a binding

allocation rule, whether provincial and lawfully devolved local discretion remains intact, and whether unequal administrative capacity distorts the data. These examples explain the test. They do not claim that either system is operational.

1.2.7 Constitutional requirements for national systems

Pakistan's Constitution requires AI governance to begin with attributed functions. Defined FLL entries may support communications, defence, economic regulation, federal administration, national planning, coordinated research and inter-provincial coordination. Article 142(c) preserves provincial competence over residual sectors, while Article 140A makes local responsibilities depend on provincial law. Articles 144–150, the CCI and the Digital Nation framework provide routes for intergovernmental action. None makes technology a federal legislative head. Authority to fund, procure, host or standardise a system remains distinct from authority to decide a case.

A constitutionally adequate national system must preserve sector mandates, custodians' responsibility for data, provincial participation, lawful local discretion and accessible remedies. Federal fit fails when technical control moves upwards but accountability stays below, when financial influence is treated as legal competence, or when an affected person cannot identify who must correct an error. Knowing who built the AI is insufficient. We need to know which institution controls each consequential step, under which law, and to whom it remains answerable.

1.3 Institutional Mandates, Oversight Relationships, and Review Routes

AI may pass through many institutions before it produces an administrative outcome. A ministry may propose the system, a technical body design it, a line agency procure it and a vendor supply it. A professional may use it under a sector regulator's supervision, while an auditor examines it, an ombuds institution hears a complaint and a court reviews the resulting act. The legal inquiry concerns each body's authority over the underlying function, its access to evidence, the action it may take and the route for reviewing its decisions. An ‘AI’ label tells us little about these matters.

This approach guards against two errors. The absence of ‘AI’ from an enabling law does not establish a legal vacuum. Powers concerning licensing, records, inspection, complaints, professional discipline, expenditure, maladministration, rights and judicial review may reach algorithmic conduct where the statutory trigger is met. Equally, a broad digital-policy or technical mandate does not confer universal regulatory jurisdiction. National coordination and ownership of technology cannot transfer a sector regulator's powers or a line department's responsibility for an individual decision.

Functions, Powers, and Institutional Outputs

An institutional map must separate functions often grouped together as 'oversight'. Authorisation provides the legal basis for a programme or decision. Deployment includes acquisition, configuration, integration, operation and control of changes. Supervision checks continuing compliance within a regulated relationship. Investigation compels or evaluates evidence of a possible breach. Audit tests expenditure, controls, compliance, performance or information systems against defined criteria. Review may reconsider the merits or examine legality. A remedy responds in a legally recognised form, whether by changing a decision, restraining conduct, imposing a sanction, compensating injury or another authorised measure.

These functions produce different results. Merits reconsideration asks what the outcome should be and may revisit facts, judgment or discretion where the governing scheme permits. Legality review asks whether the decision-maker stayed within jurisdiction, followed required procedures, respected rights, considered lawful matters and gave adequate reasons. It does not automatically remake the administrative decision. An audit may uncover waste, weak controls or non-compliance without giving an individual relief. A rights inquiry may identify a pattern and recommend correction without licensing a product or determining a statutory appeal. Professional discipline concerns a practitioner's conduct; complaint handling addresses a grievance within the institution's defined jurisdiction. A systemic recommendation seeks organisational or legislative change and may lack an order's coercive force. Recording only that an institution is 'involved' hides the differences that determine what it can actually do.

Formal competence and operational capacity also need separate entries. Competence comes from the Constitution, an Act, valid subordinate legislation or a power lawfully conferred under it. Capacity depends on suitable staff, funding, procedures, independence, technical expertise, secure facilities and actual access to records, data, logs, models and vendors. A statutory power to demand documents does not prove an ability to reproduce an output or examine a complex model. Technical expertise, in turn, gives no authority to compel evidence or bind an affected person.

Parliament, Executives, and Responsible Administration

Parliament enacts federal statutory authority, appropriates public money, scrutinises administration through questions and committees, and receives public audit through the Public Accounts Committee (PAC). Each function has limits. Funding an automated system does not authorise a new coercive power. A committee may expose a defect without becoming the administrator, regulator or appellate tribunal. Legislative and executive competence remain subject to the Constitution, including Articles 97, 137 and 141–143 and the Fourth Schedule. Federal executive power follows federal competence;

provincial executive power follows provincial competence, subject to the Constitution and competent federal law.

Article 99(3) and the Rules of Business, 1973 allocate federal portfolios and responsibility among Divisions, Ministers and Secretaries. Rule 3 and Schedule II place federal IT and telecommunications, including e-government and software standardisation, with the Information Technology and Telecommunication Division. This identifies the responsible portfolio. It does not enlarge constitutional competence or empower the Division to decide a health, education, banking, policing or provincial-service case. A department deploying a system must identify the law authorising the substantive decision and the officer responsible for it. Its records should distinguish policy coordination, budget control, procurement and technical hosting from decision-making power.

National Digital Coordination and Implementation

The Digital Nation Pakistan Act, 2025 establishes the National Digital Commission (NDC) and Pakistan Digital Authority (PDA). The Prime Minister chairs the NDC, which includes all four Chief Ministers. It approves the National Digital Masterplan, coordinates governmental entities, issues alignment directives and considers referrals of non-compliance (ss. 3–5, 11–12). This membership creates a structure for intergovernmental coordination. The constitutional allocation of functions still governs its work.

PDA undertakes planning, coordination, standard-setting, assigned compliance and oversight of digital service journeys. It develops and monitors the Masterplan and reviews projects. Subject to section 28, section 8(b) directs it to issue and enforce implementation regulations, guidelines and standards. Section 8(g) authorises oversight, standards and compliance in NDC-assigned fields, including AI, subject to existing law and the requirement to avoid overlap. Section 8(n) concerns ownership and oversight of digital citizen journeys and public-sector service delivery; section 8(h) preserves data custodians' control and responsibility. These powers confer no universal licensing, adjudicative or merits-review jurisdiction. Section 15 concerns accounts and financial audit, while section 16 requires annual reporting of activities and KPIs. Neither provision creates an independent model audit. Section 29 raises questions about constitutional review under Articles 175E–175F, 199 and 202A. Its wording alone cannot establish that technical systems are unreviewable.

The National Information Technology Board Act, 2022 establishes NITB principally as a federal e-government implementation and technical body. Section 8, particularly paragraphs (a), (c)–(h), (k)–(m) and (q)–(t), authorises technical guidance, redesign of business processes, software and infrastructure standards, cyber and IT-security alignment, readiness reviews, capacity building and provincial assistance. It also permits

NITB to facilitate the conception, procurement, operation and maintenance of federal e-government systems. NITB may contract for work under sections 7(b) and 11 and is subject to Auditor-General audit under section 18. It may therefore build or operate systems, advise institutions and hold technical evidence. These roles do not transfer a line institution's statutory discretion, complaint jurisdiction, duty to give reasons or responsibility to implement a remedy.

Sector Regulators and Statutory Commissions

Sector regulators reach AI through the activities within their mandates. Under the Pakistan Telecommunication (Re-organization) Act, 1996, as amended by the Establishment of Telecommunication Appellate Tribunal Act, 2024, PTA licenses and supervises telecommunications systems and services. It receives spectrum applications, refers allocation or assignment to the Frequency Allocation Board, and licenses or auctions the spectrum allocated or assigned. Its powers over user protection, complaints, information, inspection, licences, directions and sanctions may reach algorithms in regulated telecommunications services. Section 7(2) permits appeal to PTA from a delegated officer's decision within thirty days and directs PTA to decide within thirty days. Under section 7(1), a PTA decision or order alleged to contravene the Act may be appealed to the Telecommunication Appellate Tribunal within thirty days, with decision directed within ninety days. Section 7(3) contains the civil-court bar. Section 7B permits appeal from the Tribunal to the Supreme Court within sixty days. These routes remain specific to telecommunications. They create no jurisdiction over every connected system, and technical access still needs a legal basis.

The State Bank of Pakistan (SBP) exercises central-bank functions under the State Bank of Pakistan Act, 1956 and supervisory powers under the Banking Companies Ordinance, 1962 and Payment Systems and Electronic Fund Transfers Act, 2007. Within those laws, it may license or authorise institutions and systems, require returns and information, inspect books and accounts, issue directions and remedial measures, and impose statutory sanctions. Its payment-system powers include designation or authorisation, supervision, inspection, directions, revocation and penalties. These powers may reach regulated providers' automated credit, fraud or payment controls. Prudential supervision remains distinct from an individual complaint through SBP's channel or, where legally competent, the Banking Mohtasib.

The Securities and Exchange Commission of Pakistan (SECP) is constituted under the Securities and Exchange Commission of Pakistan Act, 1997. It exercises powers under that Act and the legislation it administers, including the Companies Act, 2017 and Securities Act, 2015. Registration, licensing, disclosure, reporting, inspection, investigation, access to evidence, directions, orders, authorised administrative sanctions and investor complaints may reach automated trading, underwriting, governance or

disclosure practices. The connection must be to regulated corporate or market conduct. A company's use of AI, considered in isolation, does not establish jurisdiction.

The Virtual Assets Ordinance, 2025 provided for a Pakistan Virtual Asset Regulatory Authority (PVARA) to license, regulate and supervise virtual asset service providers. Its original framework authorised inspection, information requirements, customer safeguards and enforcement, and required cooperation with financial and law-enforcement bodies (ss. 6, 9, 12). These provisions offer a dated example of sector-specific jurisdiction that may reach automated onboarding, transaction monitoring or market controls within the regulated activity. They do not create general AI jurisdiction. The Ordinance's constitutional duration and any extension must be assessed under Article 89; its original text should not be treated as proof of permanent legislative authority. Its scope, exclusions and coordination provisions also require separate examination before attributing a particular automated decision to PVARA.

The Competition Commission of Pakistan (CCP) applies the Competition Act, 2010 to prohibited agreements, abuse of dominance, deceptive marketing and unlawful concentrations. It may initiate an inquiry on its own motion or on a qualifying complaint or reference. It can summon witnesses and documents using civil-court-type powers, require information, conduct entry and search subject to statutory safeguards, issue remedial orders and impose penalties. Algorithmic pricing, ranking, recommendations or matching may be the conduct examined or evidence of it. AI use alone does not establish a competition-law theory of harm. The complainant must connect the system to a statutory prohibition.

Health, Professional, Educational, and Procurement Controls

The Drug Regulatory Authority of Pakistan (DRAP) administers the Drug Regulatory Authority of Pakistan Act, 2012, the Drugs Act, 1976 and subordinate therapeutic-goods regimes. Its remit covers product registration, establishment and manufacturing licences, standards, imports and exports, supply-chain controls, pharmacovigilance and enforcement. Authorised inspectors may examine premises, processes and records, take samples and seize material within statutory conditions. Registration or licensing may be suspended or cancelled, and offences follow the prescribed enforcement and drug-court routes. Software legally classified as a medical device, or algorithmic evidence concerning therapeutic-goods compliance, may fall within this remit. It does not extend to every clinical decision-support system, hospital-administration tool or allegation of professional negligence.

Under the Pakistan Medical and Dental Council Act, 2022 (Act IV of 2023), the Pakistan Medical and Dental Council (PMDC) registers and licenses practitioners, sets professional and ethical standards, recognises and inspects education, training and approved training institutions, and conducts discipline. It may examine improper

professional reliance on an automated tool. It does not register products, regulate hospital services generally or provide a merits appeal against a clinical outcome.

The Higher Education Commission (HEC), established under the Higher Education Commission Ordinance, 2002, formulates higher-education policy, sets recognition conditions, supports quality assurance and accreditation, evaluates institutions and recognises qualifications. It also allocates federal grants and collects sector information. Valid recognition, accreditation, quality or funding instruments may influence automated admissions, assessment, research and integrity systems. They do not establish a general merits appeal from every university decision. Provincial higher-education bodies, university statutes and each institution's legal character remain relevant. Calls for clearer healthcare AI guidance concern implementation; they do not establish an absence of ordinary statutory control (Umer et al., 2023).

Provincial healthcare commissions and consumer-protection authorities or courts provide further routes for licensing, inspection, complaints, enforcement or remedies within their territories and statutes. Their powers may reach deficient healthcare mediated by algorithms, misleading automated claims or defective services. Statutory definitions, exclusions, time limits and territorial reach determine whether they do. A healthcare commission does not thereby become a product or professional regulator, and a consumer forum does not become a public-law merits tribunal.

Procurement provides another form of control. The Public Procurement Regulatory Authority (PPRA) regulates and monitors federal procurement under the Public Procurement Regulatory Authority Ordinance, 2002, and federal procuring agencies follow the Public Procurement Rules, 2004. Record and publication requirements, pre-contract bidder complaints to agency committees, and applicable review or debarment routes concern procurement legality and supplier participation. Tenders and contracts may require documentation, testing, logs, audit access, cybersecurity, change management, incident notice, subcontractor disclosure and exit assistance. PPRA does not become the system owner, safety certifier or rights reviewer. The purchasing and deploying agency retains programme authority and responsibility for managing the contract. Provincial procurement remains governed by provincial laws and authorities.

Audit, Maladministration, Rights, and Review

The Auditor-General of Pakistan has constitutional standing under Articles 168–170 and operates under the Auditor-General (Functions, Powers and Terms and Conditions of Service) Ordinance, 2001. Audits cover federal and provincial expenditure, receipts, accountabilities and public institutions. The Auditor-General may inspect offices, books, accounts, vouchers and documents relevant to those inquiries. Under Article 171, federal audit reports are submitted to the President for laying before both Houses of Parliament, and provincial reports to the Governor for laying before the Provincial

Assembly, followed by legislative and Public Accounts Committee (PAC) scrutiny. In addition to financial review, compliance, performance and IT audits may expose unlawful expenditure, poor procurement, weak internal controls, cybersecurity defects or failure to achieve programme objectives.
These reports support accountability to the legislature. The Auditor-General does not thereby acquire power to reverse an individual official's administrative decision, initiate professional discipline or impose sectoral penalties.

Wafaqi Mohtasib, the National Ombudsman, investigates maladministration by federal entities on complaint or on his own initiative under the Establishment of the Office of Wafaqi Mohtasib (Ombudsman) Order, 1983 and the Federal Ombudsmen Institutional Reforms Act, 2013. He may compel attendance, hear testimony and obtain records and documents. Following investigation, he reports findings, recommends corrective action and examines the entity's response. Unremedied injustice or defiance may be reported or referred to the President under the governing provisions. Review and representation are governed by statute. Provincial ombudsmen exercise analogous powers within their respective statutory and territorial limits, including under the Punjab Office of the Ombudsman Act, 1997 and the Establishment of the Office of Ombudsman for the Province of Sindh Act, 1991. A provincial ombudsman may examine delayed responses, withheld information, unequal treatment of similarly situated people, mechanical reliance on a numerical rating without considering relevant factors, or mishandled complaints. Such an inquiry supplies neither certification of a system nor a general merits appeal.

Information Commissions established under federal and provincial freedom-of-information laws may enforce public agencies' record-access duties within the statutory coverage and exceptions. They cannot reconsider the merits of the official's original decision.

Service Tribunals established under Article 212 of the Constitution and/or federal or provincial service laws exercise specialised jurisdiction over matters concerning covered public employees.

An internal review mechanism may correct facts or reconsider the merits only where its enabling law confers that power. Each of these routes should be distinguished from judicial review for constitutional legality.

Under the National Commission for Human Rights Act, 2012, NCHR may inquire on petition or suo motu into human-rights violations, abetment and public-servant negligence. It may apply to join proceedings, visit detention facilities, review laws and safeguards, conduct research and report recommendations (s. 9(a)–(l)). Section 12 governs inquiries into complaints. Section 13(1)(a)–(f) confers civil-court powers to summon witnesses, compel documents, receive affidavits, requisition public records and commission evidence; section 13(2) concerns compelled information, subject to legal

privilege. Under section 18, NCHR may recommend prosecution or other action and immediate interim relief, transmit and publish its inquiry report, and require a governmental response. These powers could support a detailed evidential inquiry into discriminatory scoring, surveillance, exclusion or denial of a remedy. They do not provide licensing, sectoral sanctions or binding reconsideration of merits. The Commission's rights mandate must be distinguished from general regulation of AI.

Complaint routes can correct facts or remake decisions before constitutional litigation only where the reviewer has the necessary authority, access to records and model versions, and effective remedies. Rasool and Rasool (2022) identify broader difficulties with expert evidence. Assessing a particular institution's capacity instead requires its annual reports, sanctioned posts, budgets, caseloads, inspection records and documented outcomes.

Judicial Review after the Twenty-Seventh Amendment

The Constitution (Twenty-seventh Amendment) Act, 2025 redistributed superior-court jurisdiction. Article 175E(1) gives the Federal Constitutional Court (FCC) exclusive original jurisdiction over disputes between federal and provincial governments, with judgments confined to declarations under Article 175E(2). Article 175E(3) separately confers fundamental-rights jurisdiction in matters of public importance. Article 175E(5) allows the FCC to call for a pending court case's record where it considers that a substantial question of constitutional interpretation arises. Article 175F governs constitutional appeals, including appeals by leave from Article 199 decisions under clause (1)(c). Article 202A(3) reserves Article 199 jurisdiction to a High Court's Constitutional Benches where Article 202A has been brought into force for that High Court under clause (7). Subject to Article 175F, the Supreme Court retains appellate jurisdiction under Article 185.

Calling for a pending court case under Article 175E(5) is not an administrative audit. In properly instituted proceedings, the competent court may use its constitutional and procedural powers to obtain evidence and examine jurisdiction, lawful authority, fair process, reasons and fundamental rights. It does not become a permanent system auditor or the default decision-maker on the merits. A jurisdictional bar such as section 29 of the Digital Nation Pakistan Act must be considered within this revised constitutional structure. Effective review may require the administrative record, model version, input provenance, output and override logs, validation records, incident history, procurement terms and recorded reasons. Formal judicial power does not establish that technical expertise is available or that the administrator has created and preserved a complete record.

Selective Institutional-Responsibility Matrix

This preliminary mandate map records operational capacity as ND (not publicly documented) unless current public indicators for the particular institution establish its staffing, budget, expertise, caseload, access to evidence and activity.

Table 1.1 | Selective Institutional-Responsibility Matrix

Institution	Legal basis, reach and function	Decision, investigation or enforcement	Review, reporting and coordination; capacity
Parliament / executives	Constitution; Rules of Business; within the relevant competence	Legislation, funding, rules, administration and scrutiny	Political / judicial review; budget/audit reports; coordination across government levels; ND
NDC/PDA	*Digital Nation Pakistan Act, 2025*; intergovernmental digital governance	Masterplan, standards, assigned compliance, information, notices and referrals	Oversight Committee / constitutional review; annual / financial reports; coordination across government levels; ND
NITB	NITB Act 2022; federal e-government	Technical design, standards, security and procurement / operational support	Line-agency review; AGP audit; provincial assistance; ND
PTA	Telecom Act 1996; telecommunications	Licensing, information, complaints, directions and sanctions	Internal, Tribunal and Supreme Court routes; sector coordination; ND
SBP/SECP	Banking, payments, corporate and securities statutes	Licensing, returns, inspection / inquiry, directions and penalties	Statutory / complaint routes; sector reporting and coordination; ND
PVARA	Virtual Assets Ordinance 2025; original sector framework	Licensing, supervision, inspection, complaints and enforcement	Ordinance routes and coordination; duration subject to Article 89; ND
CCP	Competition Act 2010; competition conduct	Inquiry, evidence and search, remedial orders and penalties	Statutory appeal; coordination between regulators; ND
DRAP	DRAP/Drugs Acts; therapeutic goods	Registration, licensing, inspection, seizure and enforcement	Drug-court / statutory routes; safety reporting; ND
PMDC/HEC	Professional and higher-education statutes	Discipline, training, recognition, quality, grants and information	Internal / statutory / constitutional routes; sector coordination; ND
Provincial healthcare / consumer bodies	Provincial statutes; services and claims within territorial limits	Licensing / inspection or complaints / remedies, as applicable	Provincial review and reporting routes; ND

Institution	Legal basis, reach and function	Decision, investigation or enforcement	Review, reporting and coordination; capacity
PPRA/ procuring agency	Federal procurement statutes; provincial equivalents	Monitoring, records, bidder complaints and contract enforcement	Procurement review and reporting; separate systems; ND
AGP/PAC	Constitution arts. 168–171; AGP Ordinance	Evidence for financial, compliance, performance and IT audits	Legislative scrutiny/audit reports; no merits relief; ND
Ombudsmen	Federal / provincial statutes; maladministration	Inquiry, witnesses, documents and findings	Review / representation; agency follow-up and reporting; ND
NCHR	NCHR Act ss. 9, 12–13, 18; protected rights	Inquiry, evidence, recommendations for interim relief/action, and reporting	Government response and follow-up; no merits appeal; ND
Information commissions	Federal / provincial RTI statutes	Record-access complaints and disclosure orders, subject to exemptions	Statutory review and reporting; no merits reconsideration; ND
Service/ internal reviewers	Constitution art. 212; service/sector laws	Covered service appeals or legally conferred reconsideration	Statutory appeals; merits powers depend on the scheme; ND
FCC/ High Courts/ Supreme Court	Constitution arts. 175E–F, 185, 199, 202A	Constitutional evidence and orders, legality/rights review and appeals	Constitutional appellate structure; binding judgments; ND

Overlap, Capacity, and the Correct Diagnosis

Overlapping jurisdiction may be legitimate where institutions regulate different relationships and produce different kinds of decision. It becomes harmful when two or more institutions each possess decisive power in a dispute and consequently impose conflicting duties or demand duplicate compliance. Harmful overlap also occurs when the route to redress is obscured. Fragmentation creates a different problem. Each actor sees only its own part of the process, with no lawful way to connect the system owner, data custodian, vendor, frontline decision-maker, regulator and reviewer. We must identify which problem an arrangement actually presents.

Ahmed et al. (2025) describe Pakistan's civilian cyber and AI framework as fragmented in its allocation of roles, constrained by resources and political instability, and dependent on military-connected technical expertise and donor-funded technology. These are scholarly findings. They help frame an inquiry but cannot replace evidence of the particular institution's mandate, budget and performance.

Four diagnoses call for different responses. A missing-authority gap exists where none of the institutions concerned has sufficient power to act. A mandate or coordination gap arises where powers overlap but boundaries remain unassigned or mechanisms for referring disputes are missing. A knowledge or evidence-access gap exists where a competent institution lacks the necessary datasets, documents, logs, model versions, vendor materials, validation records or subject expertise. An implementation-capacity gap arises where authority and access exist in theory but inadequate funding, staffing, skills, procedures, independence or enforcement make the authority ineffective in practice.

Only the first gap necessarily calls for additional statutory power. The others may require clearer allocation of responsibility, referral procedures, contractual duties to disclose vendor materials and recordkeeping requirements for compliance activity. Training, resources and inspection capability may also be needed, alongside independent scrutiny of the systems concerned and published complaint procedures. The response should address the identified defect rather than assume that every weakness requires a new law.

The final institutional dataset should record legal basis, territorial and sectoral reach, decision-making and evidence powers, enforcement outputs, review routes, reporting and coordination duties, and publicly documented capacity. 'Not publicly documented' must remain distinct from 'absent'. An institution may be competent without demonstrated readiness, or technically capable without coercive power. Adequacy depends on connecting lawful authorisation, attributable administration, deployment that preserves evidence, distinct forms of scrutiny and an accessible remedy. A single institution carrying an AI title does not itself establish those connections.

1.4 Governing Instruments, Accountable Delegation, and Institutional Adequacy

AI governance operates through several kinds of instrument. A high-consequence system may at once be constrained by the Constitution, authorised by statute, administered under rules, purchased by contract and operated under departmental directions. A technical standard may guide its assessment, while binding precedent governs review of the resulting act. These instruments interact without taking on the same legal status. For each, we need to ask who issued it, under what authority, whom it addresses, what legal or practical effect it has, and which forum or mechanism can enforce it. Examining source, addressee, effect and enforcement gives a more accurate account than placing everything on a fixed scale labelled 'law', 'policy' and 'guidance'.

Constitutional control, executive organisation, and precedent

The Constitution establishes supremacy and requires public action to remain lawful. Article 4(1) recognises the right of 'every citizen, wherever he may be, and of every other person for the time being within Pakistan' to legal protection and treatment in accordance with law. Article 4(2) gives this principle concrete effect. Action detrimental to life, liberty, body, reputation or property must accord with law; nobody may be prevented from doing what the law does not prohibit or compelled to do what it does not require (Constitution of the Islamic Republic of Pakistan, 1973, art. 4(2)(a)–(c)). Article 8(1)–(2) makes laws inconsistent with fundamental rights void to that extent and prohibits the State from taking away or abridging those rights by law. These limits apply to AI-enabled administration. A strategy, circular, technical standard, contract or model output cannot alone justify coercion. A predictive score cannot create an offence, a procurement specification cannot authorise surveillance, and a policy objective cannot remove an existing duty or remedy.

Statutes remain subject to the Constitution, while instruments deriving force from them remain subject to the statute. An Act may establish an institution, define its territory and functions, confer powers, impose duties, authorise licensing or inspection, and prescribe appeal and enforcement routes. Subordinate legislation binds only within the authority and form supplied by its parent Act. Policy and administrative directions may guide an existing power, but their effect depends on source, wording, addressee and consistency with superior law. Constitutional scholarship on Pakistan emphasises that administrative legitimacy must be traced to the constitutional and statutory order. An institution's assertion of authority is insufficient (Aziz, 2018; Waseem, 2006).

Articles 90 and 99 locate and organise federal executive action without granting unrestricted control over digital activity. Article 90(1) provides for federal executive authority to be exercised in the President's name by the Federal Government, comprising the Prime Minister and federal ministers and acting through the Prime Minister. Article 99(1)–(3) requires executive action to be expressed in the President's name and authorises rules on authentication, allocation and conduct of business. The Rules of Business, 1973, as amended up to 24 March 2025 (Government of Pakistan, Cabinet Division, 2025), allocate work among Divisions (r. 3), assign ministerial and secretarial responsibility (rr. 5–6), and regulate authentication, agreements and contracts (r. 7). They identify the competent Division and accountable officers. They do not enlarge legislative competence, create coercive jurisdiction or validate an act unsupported by substantive law.

In Mustafa Impex, Karachi v Government of Pakistan, PLD 2016 SC 808, the Supreme Court held that the Federal Government is the collective constitutional body of the Prime Minister and federal ministers, that the Rules of Business are mandatory, and that authentication must be distinguished from statutory authority. An AI-related

power vested in the Federal Government therefore cannot be exercised by an official, minister, technical body or vendor unless the governing law and valid executive arrangements permit it.

Judicial precedent has a constitutional effect that policy and administrative practice do not possess. Under the post-Twenty-seventh Amendment text, a Federal Constitutional Court decision binds every other court, including the Supreme Court, to the extent that it decides a legal question or rests on or states a principle of law. A Supreme Court decision has the corresponding effect on all other courts except the Federal Constitutional Court (Constitution of the Islamic Republic of Pakistan, 1973, art. 189(1)–(2)). Subject to Article 189, a High Court's decision on a question or principle of law binds its subordinate courts (art. 201). No agency manual, vendor contract or claim of technical accuracy can displace that authority. Binding principles concerning legality, reasons, evidence, fair procedure and remedies continue to apply even where no AI-specific judgment exists.

Statutory instruments under the Digital Nation Pakistan Act

The Digital Nation Pakistan Act, 2025 illustrates why an instrument's label cannot settle its effect. Section 8 assigns PDA functions in different terms. Subject to section 28, it shall issue and enforce regulations, guidelines and standards needed to implement the National Digital Masterplan, and may issue directives with Commission approval (s. 8(b)). In Commission-assigned areas, including AI and data governance, it shall conduct oversight, set standards, enforce compliance and establish the necessary frameworks and processes. It must avoid overlap with another regulator and conflict with existing law (s. 8(g)). Its data-governance arrangements must also preserve designated custodians' control and responsibility for their datasets (s. 8(h)). The Act confers powers and duties while retaining sectoral jurisdiction and identifiable responsibility for data.

Section 11 places strategy, sectoral and implementation plans, technology standards, reference architectures and governance frameworks within the National Digital Masterplan. The Masterplan provides the central framework for digital initiatives and requires National Digital Commission approval (s. 11(1)–(3)). Its preparation requires consultation with sectoral bodies, regulators, and federal and provincial entities, followed by periodic review (s. 11(4)–(5)). A masterplan, reference architecture or standard within this scheme may therefore have effects specified by the Act. That does not turn every technical document into delegated legislation or transfer a sector regulator's statutory discretion to PDA through alignment with the Masterplan.

Section 12 puts Masterplan alignment into practice. Identified entities must align relevant policies, operations and digital initiatives and prepare alignment plans. PDA monitors and reports compliance and may recommend corrective measures or sanctions. It must notify a non-compliant public entity before referring persistent non-compliance

to the Commission. These effects remain subject to the Act's provisions on non-overlap, consultation and conflict.

Section 23 protects specified officeholders, employees, consultants and advisers from proceedings for acts or omissions in good-faith performance under the Act, unless bad faith, wilful misconduct, fraud or gross negligence is proved beyond reasonable doubt. That protection does not immunise PDA as an institution, enlarge a statutory power, validate an ultra vires decision or remove institutional duties.

Sections 24–27 make further distinctions. PDA may delegate a statutory power only through an Official Gazette notification specifying the power, its conditions or limits, and the period of delegation. It may delegate to its Chairperson, a Member or an officer (Digital Nation Pakistan Act, 2025, s. 24). Rules require Federal Government approval, Gazette notification and consistency with the Act (s. 25). Regulations must also be gazetted and remain consistent with the Act and rules (s. 26). Federal Government policy directives are a separate category: they must be gazetted, respect operational autonomy and accord with the Act's framework (s. 27). Section 24 does not list vendors as permitted delegates. On its face, a contract for technical assistance cannot amount to statutory delegation of PDA's decision-making power.

Section 28 also prevents a simple hierarchy from resolving every conflict. The Act overrides inconsistent provisions, but this rule is expressly subject to other laws giving specific protection to citizens' data rights and cybersecurity. Those protections prevail and must be followed (s. 28(1)–(2)). Persistent conflicts require consultation with the relevant stakeholders and regulators and, where necessary, Commission recommendations, directives or proposals to amend legislation (s. 28(3)). Resolving a conflict therefore requires attention to the subject and statutory wording. The instrument's date or title is insufficient.

Section 29 broadly excludes the agency, court and tribunal routes and injunctions covered by its wording. Without applicable binding precedent, we cannot give it a wider constitutional effect under Articles 175E–175F, 199 and 202A, or treat it as proof that technical systems cannot be reviewed.

Policy and technical standards

The manuscript cut-off determines how policy status should be described. The Ministry of Information Technology and Telecommunication's official register records approval of the National Artificial Intelligence Policy 2025 on 31 July 2025 (Government of Pakistan, Ministry of Information Technology and Telecommunication [MoITT], n.d.-b). It is an executive policy framework for priorities, coordination, implementation planning and budgeting. Approval alone makes it neither an Act nor gazetted delegated legislation. Its proposals for human oversight, public-sector registers, impact assessment,

audit, accountability and redress may shape implementation or the lawful exercise of discretion. A coercive duty still needs a competent legal source (MoITT, 2025).

Technical standards likewise acquire effects through the legal route used to adopt them. A standard may enter a valid rule or regulation, become an authorised licence condition or a contractual performance requirement, guide internal work, or provide evidence of accepted practice. Binding force may arise from direct statutory prescription, a valid statutory power or lawful incorporation into a rule, licence, contract or internal administrative direction. Technical authorship alone cannot supply it. The legal vehicle determines who is bound and how compliance is enforced. Comparative research identifies both the value and difficulties of governing AI through standards (Tartaro, 2023). In Pakistan, records should identify the issuing body, version, scope, incorporation wording and update process. Compliance may inform supervision, procurement or reasonable-care assessments. It cannot conclusively prove legality where the Constitution, enabling statute or individual facts demand more. Given documented problems with expert evidence in Pakistan, courts and reviewers must examine the expert's method, the standard's suitability and proof of actual compliance rather than accept a vendor certificate as determinative (Rasool & Rasool, 2022).

Procurement, contracts, and licences

Procurement turns many general governance commitments into practical controls. Article 173(1) extends federal and provincial executive authority to property transactions and contracts, subject to an Act of the appropriate legislature. Under Article 173(3), contracts made in exercise of that authority must be expressed in the President's or Governor's name and executed by authorised persons in the prescribed manner. A statutory corporation may instead contract in its own name under its enabling law. The Constitution thus recognises contracting authority while preserving legislative limits and execution requirements. A purchase agreement does not confer power to determine a person's legal position.

The Public Procurement Regulatory Authority Ordinance, 2002 has particular significance for federal procurement. Its current definition covers publicly funded purchases of goods, services and works. It also covers commercial transactions in which a private party performs a procuring agency's assigned functions, uses a public asset or receives budget, revenue, fees or charges for doing so (s. 2(l)). This can reach a private platform or AI supplier participating in the operation of a public function as well as supplying software. The Ordinance separately authorises federal rules by Gazette notification and Authority regulations consistent with the Ordinance and rules (ss. 26–27). Neither a procurement manual nor a contract term may contradict that statutory order.

The Public Procurement Rules, 2004 require federal agencies to pursue fairness, transparency, efficiency, economy and value for money (rr. 3–4). Specifications must ordinarily be generic and avoid favouring a supplier; an indispensable brand reference requires justification and equivalence (r. 10). Open competitive bidding is the ordinary procurement method (r. 20). Rule 23 matters particularly for AI. Bidding documents must be precise and unambiguous, specifying contract form and conditions, specifications or performance criteria, evaluation criteria and the standards used to assess quality. Rules 35, 46 and 47 address evaluation reports, procurement records and post-award access. Rule 46 requires records and associated documents to be retained for at least five years. Rule 47 permits withholding only proprietary or public-interest information with the Authority's prior approval. These are federal requirements. The relevant provincial instruments must be examined on their own terms.

Within this framework, a contract for high-consequence AI can make governance duties enforceable between purchaser and supplier. Specifications may define permitted purposes, performance and testing criteria, data provenance, human overrides, decision and access logs, accessibility and security. Terms may require incident notification, subcontractor approval, auditor and regulator access, assistance with reasons and review, approval of model changes and preservation of evidence. They may also provide intellectual-property permissions sufficient for scrutiny, data return or verified deletion, portability, service continuity, step-in rights, allocation of liability and an orderly exit. Service levels can attach response times and remedies to these duties. The obligations need precision. Calling a system ‘accurate’, ‘ethical’ or compliant with ‘industry standards’ may establish neither a testable requirement nor adequate evidence of performance.

A contract's legal force has clear limits. It binds the parties without becoming legislation. It cannot authorise the public decision, exclude constitutional or statutory review, restrict an auditor's lawful access or transfer discretion vested in a named authority or professional. Confidentiality and intellectual-property clauses should separate legitimate commercial protection from disclosure required by law to officials, regulators, auditors, courts or affected people. A liability cap may divide loss between purchaser and supplier. It cannot remove an applicable constitutional, statutory, regulatory or procedural duty to give reasons, reconsider a decision, correct records or implement a remedy.

A licence combines different kinds of condition, but derives its binding force from the enabling statute. Under the Pakistan Telecommunication (Re-organization) Act, 1996, conditions may concern statutory compliance, inspection and information, service quality, non-discrimination and consumer protection (s. 21(4)). Section 22 governs modification. Under section 23, a contravention may lead, after written notice, to a reasoned enforcement order and specified penalties. Where a valid licence incorporates

an AI-related technical standard, the statutory licensing relationship may bind the licensee to it. The condition does not thereby bind people outside the licence or permit the regulator to exceed its Act.

Retained public responsibility and institutional adequacy

An operational chain may involve a database custodian, model developer, cloud provider, interface operator, professional user and final public decision-maker. Dividing the work does not remove responsibility. Where law assigns a public function to an institution, technical outsourcing cannot transfer it without a lawful statutory allocation. The institution must retain the control required by its governing scheme over lawful purpose, material criteria, relevant facts, reasons, records, supervision, reconsideration and remedy. Where it has discretion, it must retain authority to depart from an output. A human signature is inadequate when the official lacks time, information, technical access or power to test the result. Public responsibility can coexist with other duties: the supplier may answer for contractual performance and security, and the professional for applicable practice standards, while the institution remains answerable for the public act.

Audit shows how public agencies may be examined from outside. Articles 168–171 of the Constitution make audit mandatory, establish the Auditor-General, define the audit relationship and require reports on federal and provincial accounts to be laid before the respective legislatures. Section 8 of the Auditor-General (Functions, Powers and Terms and Conditions of Service) Ordinance, 2001 addresses whether public funds were lawfully available, spent for their stated purposes and used consistently with the governing authority. Section 14 grants the Auditor-General authority to inspect offices of vendors serving the agency and obtain accounts, books, papers, other documents and information relevant to the audit. A procuring institution should therefore include contractual provisions allowing it to obtain vendor records needed for a lawful audit. Directly compelling a vendor to produce records, however, is likely to depend on the parties' legal relationships and applicable law unless a contract or statute provides otherwise. Financial and regulatory audits remain distinct from merits review, determination of rights, professional discipline and an individual remedy. Their findings may matter to those processes, but audit cannot be treated as a substitute for them.

Formal establishment is only the beginning of institutional adequacy. We need to examine lawful authority, clear and valid delegation, an identifiable and accountable decision-maker, and a fit between the mandate and the work required. We also need to know how institutions coordinate, whether they possess legal and technical expertise, and whether they retain evidence of their activities. Transparency, independent scrutiny, access to complaints and the ability to impose sanctions or other remedies complete the inquiry. Research describes Pakistan's civilian AI and cyber governance as fragmented

and under-resourced (Ahmed et al., 2025). That assessment must remain attributed to its source and tested against each institution's current budget, staff, delegations, annual reports, audit access and practices. A gap in public documentation does not prove an absence of authority or capacity.

Institutional redesign must follow the diagnosis. Authority is missing where no competent institution has legal power to perform the proposed act; policy, standards and contracts cannot repair that defect. Overlap or fragmentation exists where powers are divided, duplicated or linked by unclear referral and information routes. The immediate tasks are then allocation and coordination. Delegation or contractual allocation is inappropriate where a vendor or model effectively exercises discretion retained by law, or where contract terms obstruct records, audit, control of changes, review or exit. Implementation capacity is weak where authority and allocation look adequate on paper but staff, expertise, funding, evidence access, procedures or remedies fail in practice. Several defects may coexist. Each needs its own response, and none establishes a need for a new AI institution before the diagnosis is made.

Chapter 1 Key Arguments

The first argument concerns definition. Legal governance should identify AI by function, consequence, scale and human control, rather than unstable descriptions such as ‘AI-powered’, ‘smart’ or ‘autonomous’. We need to ask whether a system predicts, classifies, ranks, recommends, generates, matches, detects anomalies, scores risk or initiates action. Whose interests are affected? Can an error be reversed, and can the affected person obtain reasons and a remedy? This approach separates low-consequence assistance from decision support, prioritisation, risk scoring and fully or effectively automated action. A signature does not establish meaningful human control. The person needs intelligible information, sufficient time and competence, lawful authority to depart, a usable override and a real institutional opportunity to exercise independent judgment.

Second, technology does not redistribute constitutional power. The underlying function must be traced through the FLL, residual provincial competence, territorial reach and corresponding executive authority. Federal responsibility may rest on communications, defence and specified security functions, federal institutions, banking, inter-provincial coordination, national planning or another enumerated field. Provinces remain central to residual sectors including ordinary policing and public order, health, school education, labour, land and many local services, subject to the relevant constitutional and statutory qualifications. Article 140A places local government within provincial legislation and devolution. A platform's architecture cannot determine local legal authority. The Digital Nation Pakistan Act, 2025 creates substantial planning and coordination machinery. Section 8(g) nevertheless requires PDA to avoid regulatory

overlap and conflict with existing law in Commission-assigned areas, and section 8(h) preserves data-custodian responsibility. National coordination remains constitutionally credible when shared safeguards and infrastructure support lawful sectoral and territorial mandates.

Third, institutional accountability must be documented before its absence is asserted. Parliament, executives, line departments, NDC, PDA, NITB, sector regulators, professional bodies, procurement agencies, courts, tribunals, auditors, ombuds institutions and rights commissions have different AI-related responsibilities. Legality review, merits review, financial or performance audit, maladministration inquiry, rights investigation, professional discipline, procurement challenges and systemic recommendations produce different kinds of response. They cannot be treated as equivalents. An institution's enabling statute may confer relevant authority without mentioning AI. Conversely, a broad statutory label does not demonstrate capability in that field. Each institution therefore requires separate documentation of its legal basis, territorial reach, decision-making power, access to evidence, enforcement route, reporting and coordination duties, staff, budget, expertise and demonstrated activity.

Finally, technical delegation does not transfer the underlying public responsibility to a vendor, professional, model or ‘the AI’. Databases, models, interfaces, users and final administrative acts form an operational chain. Authority, reasons, records, supervision, review and remedy must still be attributable to identifiable people and institutions. Adequacy requires lawful authority, clear delegation, a suitable mandate, coordination, legal and technical capacity, documentation, transparency, independent scrutiny, accessible complaints and enforceable remedies. The diagnosis must distinguish missing legal authority; overlapping, fragmented or poorly coordinated mandates; inappropriate delegation or contractual allocation; and weak implementation capacity. Each requires a different response. A new AI-labelled authority may be justified after that inquiry, but cannot replace it.

CHAPTER 02

Public Data Infrastructures, Procurement, and Administrative Capacity

Introduction

Public-sector artificial intelligence often faces legal scrutiny only after something has gone wrong. A person has been classified, a payment has failed, an application has been ranked, or an official has acted on an alert. But many of the choices that govern these outcomes were made earlier. An institution collected a record for a statutory purpose. Databases were linked through an identifier. A tender turned an administrative problem into technical specifications, and a contract either secured or withheld audit access. Budgets did or did not cover correction, monitoring, security, appeals, and exit. We therefore need to examine data infrastructure and procurement as substantive public-law arrangements. They are part of the legal problem from the beginning.

Pakistan's developing public digital infrastructure gives this inquiry a concrete setting. The National Database and Registration Authority (NADRA) Ordinance, 2000 establishes requirements for national identity and databases. The Digital Nation Pakistan Act, 2025 creates a regulatory body to govern digital standards while retaining responsibility with designated data custodians. NADRA's digital-identity and national-data-exchange regulations, notified in 2025, address authentication, consent, access, security, logging, and interoperable exchange. The important distinction is between an authorised technical link and a lawful administrative outcome. A CNIC match, biometric verification, data-exchange response, database flag, or model score may constitute a record or provide evidence of identity. It does not conclusively establish citizenship, entitlement, culpability, or deceit, replace professional judgement, or confer legal authority to deny a service. Several entities may create, store, transmit, interpret, and act on the same information. Their respective responsibilities must remain identifiable and traceable. Otherwise, repetition may give an inaccurate field the appearance of authority. The individual is then sent from the front-line office to a custodian, vendor, or central repository, with no organisation accepting responsibility for correction and reconsideration.

This chapter follows the life cycle of public technology. Section 2.1 examines collection, reuse, inference, quality, access, correction, retention, security, and deletion. Section 2.2 considers the creation and use of digital identity through master registries, including how record and matching errors can produce cumulative exclusion. Section 2.3 examines tender design and vendor contracting, where institutions assign responsibility for evidence, performance risk, cybersecurity, intellectual-property rights, liability, continuity, and exit. Section 2.4 sets out the professional, financial, validation, security, audit, complaint-handling, and record-keeping capacity needed to make public systems reviewable in practice. Throughout, a published control is distinguished from evidence that it operates. Where the documents leave a matter uncertain, that uncertainty is recorded.

2.1 Legal Governance of the Administrative Data Lifecycle

AI governance in public administration starts before the institution chooses a model. It starts when the institution decides to record a fact, chooses a category and identifier, assigns custody, and permits the record to pass elsewhere. Even a technically sound model can reproduce a misspelled name, obsolete address, duplicated household, incorrect family link, or historically biased enforcement label. The resulting harm need not be a 'model failure'. Unlawful collection, unauthorised reuse, an uncorrected source record, a false match, or lost meaning between databases may be responsible. Calling all these events AI errors makes it harder to identify the institution that must answer for them and the remedy the person needs.

The Ministry of Information Technology and Telecommunication published the May 2023 Personal Data Protection Bill as a draft (Government of Pakistan, Ministry of Information Technology and Telecommunication [MoITT], 2023). Its proposals must be distinguished from enacted law. They offer a framework for discussing the data lifecycle, but a draft does not give a public body coercive power, authorise an otherwise unlawful data exchange, or establish an enforceable horizontal right to erasure. Legal analysis must therefore identify the enacted authority for the particular collection, disclosure, processing operation and remedy.

The rules currently in force are spread across several sources. The Constitution imposes requirements of legality, dignity, equality, access to information, and fair process within their respective fields. Enabling statutes and valid rules authorise particular records and public functions. Right-to-information laws, electronic-record rules, archival duties, cybercrime legislation, sectoral complaint procedures, contracts, and technical controls govern parts of the lifecycle. We need to read these sources

together. A policy aspiration or an interface that allows systems to exchange data cannot supply missing legal authority.

Collection authority, purpose, and secondary use

The existence of an administrative database does not authorise its collection. Before collecting information, the institution should identify the public function, the constitutional level and statutory body responsible, and the provision that authorises or requires the record. It should also identify the necessary fields, whether people must supply them, and what follows if they refuse. Article 4 of the Constitution begins with lawful treatment. It prohibits action detrimental to life, liberty, body, reputation, or property except in accordance with law. Depending on the facts, Articles 9, 14, 19A, and 25 may also bring security of person, dignity, privacy of the home, access to information, and equality into consideration. None gives the government an independent power to collect information merely because it would be useful.

The source of authority depends on the administrative field. Identity legislation governs identity data; revenue law governs tax returns and third-party information. Benefit records depend on the law and approved programme that establish eligibility. Health and education records fall within the relevant provincial service, professional, institutional, or public-health framework. Police information depends on applicable police, criminal-procedure, investigation, or security law, while licensing information depends on the statute and rules for the particular licence. Approval of a budget, authority to procure, or a broad power to modernise administration does not necessarily permit every field, linkage, or later decision criterion. In each case, we must ask separate questions: may the institution perform the function, collect the information, disclose it, derive a new attribute, and use that attribute to impose the proposed benefit or burden?

Tax law shows what express authority can look like. Section 175A of the Income Tax Ordinance, 2001 requires specified bodies to give the Federal Board of Revenue real-time access to identified databases. Subject to section 216, it confines that information to confidential tax use. Section 175AA, inserted by the Finance Act, 2025, permits the Board to share tax-declaration information with scheduled banks for cross-matching against bank data through prescribed data-based algorithms. Banks must return the specified variance results; information received is restricted to confidential tax and related use. These provisions identify purposes, sources and processing requirements. A computational match alone still does not establish final tax liability or permit unrelated reuse.

The National Database and Registration Authority Ordinance, 2000 grants substantial powers, but also places limits on them. NADRA may modernise registration, develop databases and interfaces, and use information for identification, planning, and another purpose permitted by law. The National Data Warehouse provisions cover

feeder and user agencies, exchange, mining, cleansing, normalisation, access, continuous updating, and security. Feeder and user agencies remain responsible for timely and effective updates (NADRA Ordinance, 2000, ss. 5, 7). These provisions neither prohibit exchange altogether nor license every predictive or enforcement use. The words 'permitted by law' require a lawful basis for the later purpose and recipient. The duty to update also means that an originating agency cannot treat transfer as the end of its custodial responsibility.

Recent subordinate legislation explains some of the practical requirements. Under the NADRA (Digital Identity) Regulations, 2025, a requesting party must give notice before authentication or verification. The notice must identify the information sought, its purpose and legal basis, the retention period, who may have access, and the circumstances in which it may be shared across national boundaries. The requesting party must establish the required legal bases for processing and prove that it obtained the citizen's informed and express consent. Processing is confined to the identified purposes. Secondary uses, including selling or sharing, are specifically barred. Any disclosure of processed information must comply with applicable law, remain within the original purposes, and be governed by a data-sharing agreement. Authentication processes and the resulting access events are recorded and monitored. A party may subcontract some onboarding functions, but remains fully responsible for the accuracy and integrity of its own databases and systems (regs. 11–13, 16, 24, 27, 30). These are binding rules for a particular channel. They do not establish a general consent framework for Pakistan. Where a public service is compulsory or indispensable, obtaining consent cannot make up for the absence of a relevant statute or regulation.

The National Database and Registration Authority (National Data Exchange Layer) Regulations, 2025 further explain this division of responsibilities. Onboarding establishes whether an entity will provide or consume services involving the exchange of citizen data. These exchanges are subject to the same requirements as exchanges within a single organisation. Citizen-data exchange must have a lawful collection or exchange purpose and specified data fields. It must meet consent requirements, subject to the specified exceptions, follow data-minimisation principles, operate under a sharing agreement, and preserve transaction audit trails. An intermediary facilitating citizen-entity exchanges through NDEL may not retain exchanged entity data beyond the time needed for the processing or purpose for which it received them. Responsibility for storage remains with the original citizen-source entity or the entity onboarded at the time of exchange (NDEL Regulations, 2025 defs. 4(z); regs. 3, 7). Participation in NDEL does not give a body general authority to obtain every available record.

The institution must document the original purpose and later reuse as distinct legal events. Identification alone does not establish legal status, benefit entitlement, fraud, or security risk. A tax record collected to assess liability does not automatically become

lawful training material for unrelated policing. Nor does a clinical record automatically become an employment-risk feature, or an enrolment record an unrestricted identity-resolution dataset simply because the systems can be linked. For proposed reuse in training, prediction, ranking, fraud detection, or enforcement, the institution should identify the authorising provision, compatibility with the original function and notice, necessity and proportionality, recipient competence, permitted decision criteria, and safeguards for affected people. A material change in purpose, population, sensitivity, or consequence should trigger a fresh legal and impact assessment. The legitimacy of the original collection cannot simply be carried over.

Section 8(h) of the Digital Nation Pakistan Act, 2025 supports this approach by requiring designated data custodians to retain control and responsibility for their datasets. National coordination and exchange arrangements do not transfer the originating body's statutory function. They also do not excuse a recipient from responsibility for its decision. For consequential AI use, institutions should record the lawful purpose, authorised recipients, source provenance and impact of proposed linkage. These records should identify who can correct an error and who must reconsider the decision. Good documentation supports accountability; it cannot replace enacted authority.

Biometric, behavioural, and machine-generated information

The sensitivity of information depends on its use and consequences as well as its format. Templates derived from fingerprints, faces, irises, voices, gait, or other physical or behavioural characteristics can support authentication on a large scale. These templates are also difficult to replace after compromise and can enable persistent linkage across different settings. Behavioural data include movements, transactions, communication patterns, service use, attendance, device activity, and repeated authentication failures. Combined, these events may disclose far more than any one event reveals. Security, access controls, and purpose limits must therefore protect the template and event history as well as the original image or sample.

Inferred information raises a further difficulty. A person supplies a name, address, household relationship, or transaction. From this, a system may produce a fraud likelihood, vulnerability class, confidence in a relationship, predicted health condition, political or religious proxy, or ‘reliability’ score. These are inferences. Storing them in a database does not turn them into raw input or objective fact. Hashmi's study of genealogical computation in Pakistan shows how institutions produce administrative identity categories and assessments of reliable personhood rather than simply discover them (Hashmi, 2021). A neural network can intensify this process. It may generate correlations whose decisive features are difficult to explain and whose confidence varies between populations.

Enacted law does not presently provide a common horizontal vocabulary for all this information. The Digital Identity Regulations define and protect biometric data within their own field. The draft 2023 Bill would classify biometric and genetic data, CNIC and passport information, and health, financial, religious, political, criminal, ethnic, and related data as sensitive. Its definition of profiling would cover automated evaluation or prediction of behaviour, reliability, location, and other characteristics (MoITT, 2023, s. 2). These definitions can guide reform. They are not current general rules.

The derived attribute itself therefore needs governance. Its record should identify the source data, creator, method and version, creation date, intended use, confidence or uncertainty, validation population, recipients, and expiry or re-evaluation date. It should also say whether the entry is factual, a professional opinion, or a probabilistic inference. A person may not be able to 'correct' a prediction as they would a date of birth. They must still be able to challenge its inputs, relevance, method, or continued use and obtain human reconsideration. Generated summaries and risk profiles should be clearly marked and accompanied by provenance. Their presentation should not make them indistinguishable from verified facts.

Data quality and the legal location of error

Data quality concerns more than whether a field is accurate. It includes completeness, timeliness, uniqueness, provenance, representativeness, consistency of meaning, and fitness for the administrative purpose. An address may have been recorded correctly but become obsolete. A correct historical arrest entry may have no legal or contextual relevance to a current benefit. An accurate outcome label may reproduce an institution's past selective enforcement. Electronic form proves neither truth nor relevance to a decision. Sections 3–6 of the Electronic Transactions Ordinance, 2002 recognise electronic form and specify accessibility, integrity, and origin/destination metadata where another law requires retention. They do not make the retained assertion accurate or authorise reuse.

We need to distinguish five locations of error. A source-record error occurs when the underlying administrative entry is missing, duplicated, outdated, mistranscribed, or attached to the wrong person. The originating custodian ordinarily has the evidence and authority to amend it. A labelling error occurs when a training or validation example is annotated incorrectly, or when an administrative outcome shaped by its historical circumstances is accepted as neutral ground truth. Correcting a person's live record may leave the defective training label, dataset, and model untouched. A matching error is a false or missed link arising from names, transliteration, identifiers, household relations, biometric thresholds, or probabilistic entity resolution. Investigators need evidence of candidate matches and the thresholds used, as well as the source document.

An interoperability error occurs when information loses its meaning or currency as it moves between systems. Date formats, units, status codes, field definitions, legal categories, update times, or version mappings may differ. A system may fail to pass on a revocation or may strip away provenance. Both databases can be internally 'accurate' while the exchange is wrong. A model error occurs when adequate inputs produce an unreliable output because of the specification, poor calibration, distribution shift, overfitting, drift, or a decision threshold unsuited to the local consequence. The NIST AI Risk Management Framework treats data representativeness, suitability, provenance, model limitations, monitoring, and context as distinct matters to document (Tabassi, 2023). It provides a voluntary comparative benchmark. It is not Pakistani law.

These errors can occur together. A stale source record may be falsely matched, translated through an obsolete schema, and given an overconfident score. Locating each error identifies the body that can correct the authoritative record, the recipient that must stop relying on it, the vendor evidence needed, the downstream decisions affected, and the appropriate remedy. It can also prevent repeated referrals: the service office points to NADRA, the custodian points to the receiving department, and both point to the vendor. Research into records management in a Pakistani public institution has identified implementation, organisational, skills, and preservation difficulties in electronic-record systems (Butt, Pappel, & Oolu, 2021). A good model cannot resolve an institution's failure to establish which record is authoritative or who may amend it.

Access, correction, retention, deletion, and contestation

An effective remedy begins with finding the information. The affected person should be able to learn what is held, where it originated, and whether it was observed, supplied, matched, or inferred. They should also be able to identify its purpose and recipients, the version used in a consequential process, and the custodian and decision-maker they need to contact. The federal Right of Access to Information Act, 2017 requires public bodies to maintain and index records, computerise them where appropriate, and proactively publish decision criteria, processes, budgets, and audit or evaluation reports. Its privacy exemption concerns invasion of the privacy of an identifiable person other than the requester. Appeals lie to the Pakistan Commission on Access to Information (ss. 4–8, 16–20). Provincial right-to-information enactments provide territorially distinct routes. These laws can help a person discover information, but their exemptions and jurisdictional limits remain. They are disclosure regimes, not comprehensive codes for subject access, provenance, correction, restriction, or erasure.

NADRA offers identity-document modification and complaint channels, including its complaint-management system. Within the digital-identity regime, users must receive notice of successful and failed authentications and may access specified authentication logs (Digital Identity Regulations, 2025, regs. 24, 30). These mechanisms are useful.

They do not, however, correct every feeder-agency record, derived risk score, or linked-service decision. After a source record changes, the receiving institution still needs to locate the affected copies and outputs, mark the dispute, reconsider its decision, and communicate the correction to downstream recipients.

An effective procedure should provide four distinct remedies. Factual amendment corrects or supplements the authoritative record while preserving its audit history. Propagation sends a signed correction or revocation to identified recipients and prevents an outdated replica from silently restoring the error. Interim restriction pauses new adverse reliance where the dispute is credible and delay could make relief ineffective. Decisional reconsideration asks the body with legal decision-making power, rather than only the database custodian, to revisit the benefit, investigation, priority, licence, or other outcome. A disputed professional opinion or probabilistic inference may be preserved for evidentiary reasons while being marked as disputed, time-limited, or unfit for the proposed use.

The draft 2023 Bill would provide notice, access, correction, withdrawal, restriction of harmful processing, erasure subject to exceptions, grievance procedures, and safeguards for automated processing and human intervention (MoITT, 2023, ss. 7, 16–29). These proposals describe possible horizontal protections; their inclusion in a draft does not make them enforceable. The available route must instead be identified in the applicable identity or sector statute, RTI law, internal review, regulator or ombuds jurisdiction, and judicial review governing the act in question.

Retention requirements also depend on the source of law. The National Archives Act, 1993 covers machine-readable federal public records. It requires specified non-current records to be transferred to the National Archives and restricts destruction without the prescribed archival authority (ss. 2, 5, 7–8). The Electronic Transactions Ordinance explains how electronic records can satisfy an existing retention duty, but sets no universal period. PECA criminalises forms of unauthorised access, copying, interference, and misuse of identity information, and imposes particular preservation or retention duties within its field. Under section 16, an affected person may seek the securing, destruction, blocking, or prevention of transmission of identity information used without authorisation. This offence-specific remedy is not a general right to correct or erase lawfully held administrative data (Prevention of Electronic Crimes Act, 2016). The Digital Identity Regulations contain a more specific rule: NADRA and requesting entities must retain defined minimum authentication transaction or log data indefinitely, subject to purpose and access restrictions (regs. 25, 30). That exceptional requirement does not establish a general policy of indefinite retention.

A lawful retention schedule must distinguish the authoritative public record from operational copies, exchange caches, model-development datasets, decision and audit trails, security logs, litigation or investigation holds, and archival copies. For each, the

institution needs an identified legal basis, minimum and maximum period, review event, authorised disposal method, and evidence of destruction or transfer. Deletion must respect archival, audit, evidentiary, and statutory duties. Those duties do not justify retaining every working copy forever. Secure deletion should extend to backups, vendor environments, derived feature stores, embeddings, and feasible downstream replicas. Any exceptions and technical limits should be recorded openly.

Security is part of lawful processing, but it cannot supply legality. The NADRA Ordinance requires security, secrecy, safeguards, and confidentiality. The Digital Identity Regulations add encryption, auditable access, system testing, independent information-security audit, incident cooperation, and specified breach reporting (NADRA Ordinance, 2000, ss. 5, 7; Digital Identity Regulations, 2025, regs. 16, 27–35). These controls reduce unauthorised access and compromise. They cannot legitimise excessive collection, an irrelevant inference, or a denial based on an inaccurate record, however securely that record is stored.

A documentary data-lifecycle matrix

The following matrix identifies the minimum record needed to reconstruct a consequential system's data lifecycle. Institutions should complete it for every dataset and material transformation, rather than once for the programme as a whole. Entries must identify actual instruments, people, system identifiers, dates, and retained evidence. An assertion of ‘policy compliant’ does not answer these questions.

Table 2.1 | A documentary data-lifecycle matrix

Lifecycle event	Legal authority and purpose	Data, provenance, and sensitivity	Custody, sharing, and downstream use	Quality, correction, and contestation	Retention, security, and review evidence
Create or collect	Enabling instrument and competent institution; original purpose; compulsory or voluntary basis; necessity of each field; consequence of refusal	Field definition; source person or system; collection time; identity assurance; biometric, health, financial, behavioural, or other sensitivity	Originating custodian; collector and vendor role; authoritative-record ID; intended programmes and recipients	Entry validation and duplicate check; accessible notice; source-record correction route and service standard	Applicable schedule; encryption and access roles; device and breach controls; retained notice, version, and approval
Clean, label, transform, or infer	Authority for the transformation and intended use; compatibility or fresh approval for secondary purpose	Input snapshot and lineage; labelling protocol; derived feature or inference; confidence; method, model, and annotator	Derived-data controller; processor or subcontractor; approved environment; training, validation, or operational recipients	Sampling and label review; contextual, group, and regional checks; route to dispute facts and inferences	Retention distinct from source records; reversible or pseudonymous linkage; access logs; transformation code, approvals, and change history
Link or share	Recipient competence; disclosure provision; agreement; registered purpose and prohibited uses	Exact fields; match keys and thresholds; schema and version; timestamp; preserved provenance	Sending and receiving custodians; interface or NDEL role; onward recipients; receiving dataset and decision systems	False- and missed-match testing; semantic mapping; reconciliation; signed correction and propagation route	Transit encryption; credentials and cache period; transaction and access logs; incident and reviewer-access terms
Train, validate, or configure	Authority to use administrative data; approved task, population, outcome, and decision role	Dataset and label versions; exclusions; representativeness; geographic, linguistic, and temporal coverage; known limits	Model owner, host, vendor, subcontractors, permissions, and intended deployments	Held-out local validation; error taxonomy; subgroup and context results; unacceptable-error threshold and remediation owner	Secure environment; copy inventory and test-data disposal; vulnerability controls; validation report, configuration, and deployment decision

Lifecycle event	Legal authority and purpose	Data, provenance, and sensitivity	Custody, sharing, and downstream use	Quality, correction, and contestation	Retention, security, and review evidence
Use, decide, and notify	Decision-maker and enabling power; lawful criteria; human role and override authority	Case input snapshot; model and rule version; output, uncertainty, and limitations shown to the user	Operating department; professional user; final-decision owner; persons and systems notified	Plausibility check; override and escalation; reasons and notice; accessible complaint, interim restriction, and human reconsideration	Tamper-evident decision and access logs; least privilege; monitoring triggers; final act and review-route record
Correct and propagate	Institution empowered to amend the source and body empowered to reconsider the decision	Contested field or inference; supporting evidence; affected versions, matches, and outputs	Custodian, recipients, vendor, and complaint body; referral owner; downstream inventory	Acknowledgement and deadline; dispute marker; amendment preserving the audit history; root-cause classification and escalation	Necessary evidence preserved while reuse is restricted; secure correction; acknowledgements, revalidation decision, and remedies
Retain, archive, delete, or decommission	Statutory or sectoral schedule; archival approval; audit, investigation, or litigation hold; disposal authority	Record class, copies, backups, logs, features, embeddings, derived data, and sensitivity	Custodian, archive, vendor, and replica holders; exit duties; systems and recipients to be disconnected	Necessity review; expiry of inferences; protection of unresolved disputes; verification sample	Minimum and maximum periods; deletion or archival method; key destruction; certificate; audit trail and officer sign-off

The matrix makes lifecycle principles usable in procurement, audit, complaints, incident investigation, and judicial review. It also shows where the documents leave questions unanswered. If the institution cannot identify the source authority, authoritative record, recipient, version, retention rule, or person responsible for correction, it should record that absence as a governance finding. It cannot assume that the platform, vendor, or 'AI' holds the answer.

2.2 Digital Identity, Registries, Interoperability, and Administrative Exclusion

Digital identity provides basic infrastructure for administration. It can establish whether a claimant is the person associated with a credential, link records, and reduce repeated demands for documents. Without the governing sector law, it cannot decide citizenship, entitlement, culpability, creditworthiness, medical priority, or educational status. This distinction matters particularly in Pakistan. The Computerised National Identity Card (CNIC) provides access to many public and private transactions, while the National Database and Registration Authority (NADRA) maintains national databases and supplies verification and exchange services. A common identifier makes reuse efficient. If the reused information contains an identity or matching error, that same efficiency can spread exclusion across services.

The legally relevant system extends well beyond the card or authentication model. It includes the office that creates a birth, death, marriage, residence, disability, or household record, the feeder agency that updates the registry, and NADRA's identity and exchange services. It also includes the service body applying an eligibility rule, the interface through which an official receives a match or flag, and the institutions able to correct the record and remedy the outcome. For interoperability to be lawful and reliable, every link must have identifiable authority, custody, matching rules, security, audit evidence, correction duties, and decision-making responsibility.

Foundational identity, authentication, and substantive authority

The National Database and Registration Authority Ordinance, 2000 gives NADRA an extensive but defined registration mandate. Sections 5 and 7 authorise modern registration arrangements, multipurpose databases, data warehouses, networks and interfaces, a Citizen Database and National Data Warehouse (NDW), links with user and feeder agencies, and specified data sharing, cleansing, and normalisation. Under section 7(3), each feeder agency is responsible for continuously, promptly, and effectively updating its database and maintaining its link. Sections 8–10 govern registration and the issue of identity documents. Section 19(4) makes the Computerized National Identity Card (CNIC) proof of identity as established by its contents

(NADRA Ordinance, 2000, ss. 5, 7–10, 19). NADRA is therefore an identity registrar and infrastructure operator. It does not become the substantive administrator of every service that uses a CNIC.

Identity data contain more than a name and fingerprint. Hashmi's study of NADRA verification explains how digitised family relationships and documentary genealogies can serve as administrative evidence of reliability and belonging (Hashmi, 2021). A disputed family link is thus a data-quality problem. When an institution uses it to question a person's status, it also becomes a classification with potentially serious consequences.

Four distinctions are necessary. Citizenship is a legal status under the Pakistan Citizenship Act, 1951. A successful biometric comparison does not create it, and the Ordinance itself defines a citizen by reference to citizenship law. A CNIC provides evidence of identity, but does not establish that its holder meets a benefit threshold, has a professional licence, owes tax, poses a security risk, or qualifies for particular treatment. Authentication, in turn, asks whether the credentials presented match the identity record. The NADRA (Digital Identity) Regulations, 2025 define it through credentials such as a PIN, token, signature, or biometric, and provide for a yes/no response or specified e-KYC data. Finally, the sector institution must apply its own lawful criteria and make the decision. A negative authentication response may call for further verification. Without more, it cannot lawfully justify denying the underlying service.

The Digital Nation Pakistan Act, 2025 maintains this separation. It defines Digital ID as a legal, secure, and verifiable credential for an eligible individual, developed, issued, and managed by NADRA under the 2000 Ordinance. Digital public infrastructure includes identity and exchange systems, which the Pakistan Digital Authority (PDA) must develop. Section 8(h), however, requires designated custodians to retain control and responsibility for their datasets (Digital Nation Pakistan Act, 2025, ss. 2(i), 2(m), 8(h), 8(k)). Coordinating infrastructure does not silently transfer a health, education, police, revenue, or social-protection decision to the identity provider.

Courts have enforced this boundary where identity infrastructure was itself used coercively. In Askari Bank Ltd. v A.H. International (Pvt) Ltd. and others (Sindh High Court, 23 September 2025), the Court recalled an order blocking judgment debtors' CNICs during execution of a banking decree. The applicable Civil Procedure Code and Financial Institutions (Recovery of Finances) Ordinance, 2001 supplied no such power in Sindh. Reliance on a provision from another province was misplaced, and an incorrect address had also prevented proper notice. The Court directed NADRA to unblock the CNICs. This decision concerns identity restrictions rather than an AI model. Its relevance here is an administrative analogy: a technical means of exerting pressure does not enlarge a body's powers, and an institution must use the procedures and remedies available under the law that governs its own function.

Master repositories and federated exchange

Pakistan's current instruments provide two related arrangements with different legal foundations. The NADRA (National Data Repository) Regulations, 2024, gazetted in March 2025, treat the NDW as a centralised, multipurpose National Data Repository (NDR). Feeder agencies must maintain and update their databases to NADRA standards. An entity seeking access under section 7(i) of the Ordinance must explain its justification and purpose. NADRA then assesses necessity, relevance, and legal compliance. An approved Data Sharing Agreement must define purpose, scope, limits, security, confidentiality, retention, and responsibilities. The regulations prohibit uses outside that agreement and require periodic reporting and secure, irreversible disposal when the authorised period ends. They also state that sharing under section 7(i) does not require explicit data-subject consent, while retaining requirements of responsible and lawful treatment (NDR Regulations, 2024, regs. 12, 14–19).

The NADRA (National Data Exchange Layer) Regulations, 2025 govern federated exchange. NDEL allows standardised exchange through secure interfaces while the originating entity continues to own and manage its data. An intermediary may retain the data only for processing. Onboarded federal, provincial, public, and private entities require NADRA approval, service specifications, public-key credentials, and logged interfaces. Ordinary exchange requires a lawful and legitimate purpose, explicit consent recorded cryptographically, minimisation, security controls, local hosting, incident response, and audit. Secondary sale or resharing is prohibited. An onboarded entity must notify NADRA of a breach within twelve hours. NADRA retains NDEL transaction logs indefinitely, while onboarded entities must keep immutable logs for five years. The entity remains responsible for the security and accuracy of its NDEL operations, including subcontracted activity (NDEL Regulations, 2025, ch. II, regs. 1–17).

Digital-ID rules also place duties on entities dealing directly with citizens. A requesting entity must give a consent notice specifying the data sought, purpose, duration, recipients, and whether provision is mandatory or voluntary. It must prove consent and may not sell or reuse the data. Users are to receive information about each authentication, including the requesting entity, time, and purpose, and may access authentication logs. The rules require role-based access, encryption, security assessment, audit, backup authentication, breach reporting, reasonable accessibility measures, and Urdu, English and/or regional-language support (Digital Identity Regulations, 2025, regs. 9–16, 22–30, 34–35).

These instruments do not establish universal interoperability of government data. NDR permits centralised holding and statutory sharing under an agreement without express consent. NDEL governs exchange in which originators keep their data and ordinary sharing depends on consent. A proponent must identify the actual route, dataset, legal power, and agreement being used. Consent also leaves public-law questions

unresolved where a person must surrender data to obtain an essential service and cannot realistically refuse. Collecting and deciding bodies still need statutory authority, necessity, and proportionality. Equally, a statutory route without consent does not remove purpose limits, security requirements, constitutional legality, or the duty to make a fair decision.

Custodial responsibility needs practical controls. Each exchange should identify its authoritative source, field definitions and dates, and whether the result is an exact identifier match, a probabilistic link, or a biometric comparison. It should record thresholds, possible false matches or non-matches, transformations, recipients, and later corrections. NDEL security and transaction logs establish that an exchange occurred. They may not explain why a service agency denied a claim, what it inferred, or whether an official could override the result. The decision-maker must therefore retain the joined record, operative rule, version, reasons, notice, and reconsideration history. Existing regulations impose important security and logging duties. They do not establish a universal procedure through which individuals can correct every sector registry. The route still depends on the source custodian, NADRA's identity procedures, and the applicable service law.

How a single error becomes cumulative exclusion

A linked administrative process can fail before it uses any predictive model. A birth or marriage may be omitted or mistranscribed, a name transliterated differently, or a death or address left out of date. Two people may be wrongly linked, a fingerprint may fail to match, or an accurate identity record may be joined to an outdated household, disability, property, or employment record. The receiving system may then read ‘not matched’ as ‘not eligible’. The first describes the quality of evidence. The second is a legal conclusion. When services reuse the same identifier, a disputed record may appear independently confirmed across benefits, payments, and other services, multiplying its consequences.

The National Socio-Economic Registry (NSER) provides a documented Pakistani example that does not depend on speculation about AI. The Benazir Income Support Programme (BISP) describes NSER as a national household socioeconomic repository. It cross-validates roster fields, including name, CNIC, marital status, gender, and disability status, with NADRA, and uses proxy-means and exclusion filters to target programmes. It has moved from periodic door-to-door enumeration towards a dynamic registry using tehsil desks and NADRA validation (BISP, n.d.-c). NADRA answers the identity or roster question; BISP is responsible for the socioeconomic record and benefit process. A mismatch, outdated household composition, targeting rule, or biometric payment failure involves a different cause and may require a different institution to correct it.

An independent World Bank assessment found that NSER covered about 84 per cent of the population. Coverage was substantial, but more than 2.2 million households in

the bottom 40 per cent were missing. The estimated probability of exclusion was about 26 per cent in Balochistan and 11 per cent in Khyber Pakhtunkhwa. Almost half of unregistered survey participants did not know how to register, and only 27 per cent knew how to contact BISP (Majoka et al., 2024). These findings concern access to the registry. They do not prove that a particular automated decision was unlawful. They do show why a master registry cannot be treated as a census of everyone entitled to assistance. 'No record found' requires outreach and verification, rather than an automatic adverse inference.

Differentiated ability to authenticate and correct

Unequal digital access affects whether notice, consent, and remedies work in practice. The Pakistan Bureau of Statistics (PBS) Pakistan Social and Living Standards Measurement Survey 2019–20, published in 2021, reported household internet access at 33 per cent nationally, with 48 per cent in urban and 23 per cent in rural areas. Among people aged ten and over, 19 per cent reported using the internet in the preceding three months: 31 per cent in urban and 12 per cent in rural areas. Individual mobile ownership stood at 45 per cent, with reported rates of 65 per cent for males and 25 per cent for females (PBS, 2021, Summary of Key Indicators: ICT). These figures describe the survey period and establish a historical pattern of unequal access. A connected household does not necessarily give each member personal control. A woman using a shared handset may count as 'connected' while being unable to receive a confidential notice, preserve an OTP, inspect a transaction log, or appeal without assistance.

Distance and poverty impose further costs: travel, lost wages, and the expense of obtaining documents. The NSER assessment found that dynamic registration improved inclusion, but desk-based registration could leave remote households underserved. It recommended targeted outreach and complementary door-to-door registration, particularly in Balochistan (Majoka et al., 2024). A correction process available only through a portal would reproduce this selection problem. Effective provision requires assisted and offline channels, mobile registration, accessible local offices, receipts and tracking numbers, fee waivers where legally permitted, and a rule preventing disputed records from spreading while correction remains pending.

Biometrics can increase personal control while also creating risks of exclusion. BISP introduced fingerprint verification from 2017, materially increasing the probability that women collected and controlled their own cash. The evaluation found no average reduction in successful collection during the rollout studied. It nevertheless identified difficulties for people whose fingerprints are hard to read, including older people and manual labourers, alongside device, application, and connectivity failures. It called for accessible exception handling and grievance redress (Clark et al., 2022). The legal response should account for both findings. Where failure is foreseeable, no single

authentication method may become an unreviewable condition of access. The 2025 Digital Identity Regulations require backup authentication and accommodations, and define biometrics broadly enough to include facial and iris methods. These are published requirements. They do not prove that every downstream payment point can presently provide the fallback.

Language and disability require attention throughout the process. A multilingual application offers limited help if notices, call-centre scripts, affidavits, or appeal orders remain inaccessible. The Digital Identity Regulations require reasonable efforts by NADRA, accommodations and assistive support from integrating entities, plain-language notices, and language choice. NADRA also publishes an inclusive-registration programme and mobile services. Evidence that these channels exist does not establish equal geographical coverage, accessible staffing, or effective resolution. Service-level data and independent audit must answer those questions.

Migration raises a related difficulty of classification. Through the government–United Nations High Commissioner for Refugees (UNHCR) Documentation Renewal and Information Verification Exercise, NADRA verified and issued biometric Proof of Registration (PoR) smartcards to approximately 1.3 million registered Afghan refugees (UNHCR, 2022). A PoR card provides meaningful identification. It is not a CNIC and does not establish Pakistani citizenship. People moving within Pakistan may also hold valid identity documents while their address, family, or service records remain outdated. Systems must distinguish authentication failure from the absence of a particular document, immigration or citizenship status, residence criteria, and the sector entitlement actually under consideration.

Responsibility, correction, and remedy across the chain

Responsibility should follow the legal mandate and the ability to correct the problem. A local or sector office that creates a civil or service record must attest to its provenance and correct source errors. Under section 7(3) of the NADRA Ordinance, the feeder agency must transmit updates. NADRA answers for the citizen database, identity credential, authentication service, and exchange controls within its mandate. The Digital Nation Act and NDEL Regulations leave control and responsibility for a dataset with its originating custodian, while the onboarded entity remains responsible for exchange operations and subcontractors. The service institution must identify the substantive rule, decide entitlement, give reasons, and reconsider the case. It cannot assign those duties to NADRA, NDEL, or a vendor.

Correction must reach the systems that used the error. A workable protocol should identify the authoritative source and give the person a readable account of the disputed fields and their provenance. It should accept evidence through assisted and offline channels, suspend avoidable adverse reuse, and record the correction and effective date.

Every authorised downstream recipient should be notified, affected eligibility or risk processes rerun, and retrospective payment, restoration, or other relief provided where the error caused loss. Authentication logs can help people discover which entity accessed their identity data. The service body must still supply its decision record and reasoning. Repeated referrals and claims of vendor confidentiality do not remedy the harm.

NADRA publishes complaint channels through its 1777 call centre and online support. Section 18 of the Ordinance also provides notice and an appeal to the Federal Government for specified cancellation, impounding, or confiscation decisions. Sector programmes maintain their own complaint arrangements. These routes should use a common case identifier and a duty to transfer cases, so that individuals do not have to diagnose how the systems connect. Constitutional review remains available for public action taken without authority or due process, as the CNIC cases show. Ordinary correction should nevertheless occur before exclusion spreads and without requiring High Court litigation.

The following pathway is an analytical reconstruction. It does not claim that every listed system is technically linked in Pakistan.

Cumulative-exclusion pathway

Table 2.2 | Cumulative-exclusion pathway

Stage	Typical failure	Institution with primary control	Minimum safeguard and corrective route
1. Record creation	Omitted event, mistranscription, inaccessible registration	Civil/sector record creator and feeder agency	Source receipt; provenance; assisted correction; timely feeder update
2. Identity record	Incorrect family link, duplicate, stale attribute, disputed status	NADRA for identity record; competent status authority for legal status	Specific notice; evidence access; reasoned correction; section 18 safeguards where applicable
3. Authentication or matching	False non-match, false match, biometric or connectivity failure	NADRA and the requesting / onboarded entity	Failure code; alternative modality; manual verification; transaction log
4. Registry linkage	Wrong household/person join or stale imported field	Originating custodian and receiving registry	Match method and threshold; source/date display; correction propagation
5. Flag or eligibility calculation	Missing information treated as adverse evidence; opaque exclusion filter	Sector programme or decision authority	Published lawful criterion; validation; human reconsideration; no automatic inference from non-match

Stage	Typical failure	Institution with primary control	Minimum safeguard and corrective route
6. Notice and attempted correction	No reasons, digital-only notice, referral loop	Decision authority and complaint handler	Accessible multilingual notice; one case number; transfer duty; pause and time limit
7. Downstream remedy	Corrected source but denials persist elsewhere	Each recipient, coordinated by source custodian and programme owner	Recipient register; rerun; restoration/back payment; audit confirmation and escalation

2.3 Procurement as a System of Technology Governance

Public procurement helps put AI governance into practice. Before a system classifies an applicant, ranks a case, detects an anomaly, recommends enforcement, or generates material for an official, procurement will often have settled crucial questions. What problem will it address? Which data may it use? How freely may the supplier change it, what evidence can the institution obtain, and when can a person intervene? Procurement also affects whether the institution can end the arrangement in practice. A weak tender can leave a technically capable model beyond meaningful review. A carefully drafted contract, however, cannot supply missing legal authority for the public function.

AI procurement in Pakistan must account for both points. Procurement law governs public expenditure and the selection and management of suppliers. The Constitution, enabling statutes, sector law, administrative law, information-access rules, audit powers, and applicable rights continue to govern the system's purpose and use. The contract converts these public-law and operational requirements into obligations enforceable against the supplier. It does not create coercive governmental power or transfer to a vendor discretion that the law places with a public institution or professional decision-maker. Comparative scholarship accordingly treats procurement as an important means of ex ante AI governance, while stressing its connection to the substantive law governing deployment (Coglianese, 2024; Vázquez Matilla, 2025).

The federal procurement framework and its limits

Under Article 173(1) of the Constitution, federal and provincial executive authority extends to making contracts, subject to legislation by the competent legislature. Article 173(3) requires a contract made in exercise of that authority to be expressed in the name of the President or Governor and executed on that office-holder's behalf by a duly authorised person in the prescribed manner. A statutory corporation contracts in its own name under its governing law. These formalities identify the state entity entering the contract. They do not establish whether it may collect particular data, determine eligibility, investigate a person, or automate an administrative act. The institution must

answer those questions from its substantive mandate before procurement begins (Constitution of the Islamic Republic of Pakistan, 1973, art. 173).

The Public Procurement Regulatory Authority Ordinance, 2002 and Public Procurement Rules, 2004 provide the general federal framework. Following the 2023 amendment, 'public procurement' covers publicly financed acquisition and certain arrangements in which a private party performs a procuring agency's assigned function, uses a public asset, or receives a benefit based on a budget allocation, revenue, fee, or charge (Public Procurement Regulatory Authority Ordinance, 2002, s. 2(l)). Where the statutory conditions are met, this definition may cover a firm operating a platform or part of a public service, as well as one supplying software. PPRA may monitor procurement law, practice, quality arrangements, and agency performance, and obtain procurement information (ibid., ss. 5, 16). Those powers do not make it the substantive regulator or deploying authority for every AI system purchased.

This chapter uses PPRA's live official consolidation, labelled 'as amended up to date'. It records amendments to the Public Procurement Rules, 2004 through S.R.O. 763(I)/2025 of 15 April 2025. References elsewhere to proposed or differently titled replacement rules do not establish that those rules have commenced. Gazette status requires verification. The Rules govern federal procuring agencies. They do not automatically govern provincial or local procurement, for which the applicable instruments must be examined separately.

Several requirements in the Rules support responsible technology acquisition. Procurement must be fair, transparent, efficient, economical, and offer value for money. That concept combines price with quality, timeliness, reliability, after-sales service, upgradeability, source, and whole-life cost (Public Procurement Rules, 2004, rr. 2(l), 4). Planning must realistically assess requirements, resources, timing, and future benefits, and procurement requires competent approval (rr. 8, 11). Open bidding is the principal method (r. 20). Specifications must ordinarily be generic and avoid favouring a supplier. An indispensable brand reference needs written justification, the words 'or equivalent', and defined equivalence (r. 10). Direct contracting is limited to the conditions in rule 42. Describing AI as proprietary does not, by itself, justify avoiding competition.

Rule 23 connects general procurement duties to system design. Bidding documents must be precise and unambiguous. Among other things, they must state the contract form and conditions, specifications or performance criteria, qualification and evaluation criteria, delivery schedule, and quality-assessment standards. Evaluation criteria must be unambiguous and included in those documents. Once bids are opened, only the disclosed criteria may be used; failure to provide unambiguous criteria constitutes mis-procurement (rr. 29–30). For complex work, a single-stage two-envelope procedure can allow technical assessment before price evaluation. Two-stage procedures can refine technically complex requirements on equal terms (rr. 36–37). These rules allow agencies

to demand AI-specific evidence, but do not themselves prescribe training-data provenance, subgroup validation, explainability, model logs, or human override. Unless another applicable law or standard requires these safeguards, the agency must specify them in advance as proportionate technical, evaluation, and contractual requirements.

PPRA's standard information-systems and cloud-services documents provide useful general and special conditions, requirements, and arrangements for acceptance, service levels, security, and contract management. They do not certify AI assurance. The Pakistan Cloud First Policy 2022 and the notified National Standard Bidding Document for Procurement of Cloud Services (S.R.O. 730(I)/2024) together address service levels, security, audit, incident handling, the data lifecycle, and contract management. The Policy is an executive-policy instrument. It does not independently authorise data use or administrative decisions. The notified bidding document operates under rule 23(4), but the agency must still lawfully complete the requirements and special conditions for its particular procurement.

Problem definition before specification

Before procuring AI, an institution must establish whether acquisition addresses a problem lawfully and justifiably. Its business case should identify the statutory or executive function, responsible institution, operational failure or need, and measurable public outcome. 'Modernisation', 'smart government', and 'AI-enabled efficiency' do not describe what a system must do. The planning discipline in rule 8 and competent-approval requirement in rule 11 support a recorded assessment of existing processes, non-AI alternatives, internal development, shared services, and procedural reform. A vendor's ability to demonstrate a model does not establish a need to buy it. If the underlying problem is inaccurate source records, insufficient staff, or an inaccessible complaint process, prediction may simply automate its effects.

The business case should explain the proposed function: generation, matching, classification, ranking, prediction, scoring, detection, recommendation, or automated action. It should identify the people affected directly and indirectly, the stage at which the output enters a decision, and the authority responsible for the final act. It should account for children, people with disabilities, linguistic minorities, migrants, people without reliable documents, and rural or low-connectivity users. Procurement cannot be designed solely around users who can authenticate, read a digital notice, and pursue an online appeal.

The consequences of use determine the error standard needed. An assistant drafting internal text that staff can review differs from a system that prioritises police attention, flags benefit claimants, matches biometric identities, or recommends licence denials. A tender should specify the false positives and false negatives that matter, how they will be measured, the period, subgroups and locations covered, and the consequences of each

error. Average 'accuracy' alone is inadequate where errors fall unevenly or an apparently correct prediction uses a legally irrelevant proxy. Generative systems require assessment of unsupported statements, omitted qualifications, source attribution, instruction-following, harmful content, and uncertainty. A classification-accuracy measure cannot answer all these questions. Acceptance thresholds should reflect reversibility and the available remedy. Placing a nominal human reviewer after a score that is difficult to challenge does not justify a low threshold.

Procurement must provide for human control. Requirements should state what users see, which information and reasons they receive, how much time they have, who can depart from an output, whether departure needs approval, how disagreement is recorded, and how they can examine the original evidence. Interface defaults, workflow deadlines, bulk processing, performance targets, and a lack of alternatives must not make a recommendation compulsory in practice. For high-consequence uses, acceptance testing should establish whether an authorised official can pause, override, escalate, and reconstruct an output without vendor intervention. Training, staffing, and time form part of system performance. They cannot simply be assumed.

These requirements should be tested through evaluation criteria. Rules 23 and 29–30 allow performance criteria and require advance disclosure and consistent application. A complex procurement may first set mandatory conditions for lawful architecture, security, validation, audit access, and exportability, then weigh quality and whole-life cost. The 'most advantageous bid' may combine cost, quality, and qualification (rr. 2(h), 38). A pilot also needs a bounded purpose and population, lawful data access, and success and stopping criteria. Expansion should not be presumed. If a pilot affects legal interests, it remains a public deployment.

Vendor, model, data, and supply-chain diligence

The Rules allow prequalification for services and expensive or technically complex equipment by reference to relevant experience, personnel and equipment, financial position, managerial capability, and other consistent factors. Agencies must disclose documentary requirements. They may make written inquiries where credible concerns arise about professional, technical, financial, legal, or managerial competence, and must disqualify bidders who provide materially false, inaccurate, or incomplete qualification information (Public Procurement Rules, 2004, rr. 15–18). Rule 19 and the 'Mechanism for Blacklisting and Debarment of Bidders or Contractors Regulations, 2024' provide a further mechanism for established corrupt or fraudulent practices and specified performance failures (S.R.O. 460(I)/2024, regs. 5–11). Agencies must follow applicable beneficial-ownership and integrity-pact requirements and should check the public debarment register.

The following AI inquiries are proposed applications of these provisions. The generic Rules do not enumerate them. A bidder should identify the legal entity responsible for each component, the developer and model version, and whether the model is bespoke, open-weight, licensed, or supplied through an API. It should disclose material dependencies, hosting, and every subcontractor or subprocessor able to access data, alter behaviour, review outputs, or interrupt service. Disclosure should also cover relevant litigation, enforcement, debarment, incidents, beneficial ownership, and conflicts involving advisers or evaluators.

Claims about training and evaluation need evidence proportionate to access and risk. The purchaser should examine data provenance and legal availability, collection period and population, labelling methods and reliability, known gaps, duplicates, synthetic data, and proxy variables. It should also examine preprocessing and exclusion rules, separation of training and test material, external validation, earlier deployment limitations, and performance across relevant languages, sexes, ages, disabilities, regions, and other legally or operationally material groups. A foundation-model purchaser may not receive the full pretraining dataset. It can still require a documented account of provenance practices, relevant evaluations, prohibited uses, model and safety documentation, retrieval sources, and data-use settings. It can also require evidence that the provider will not train on institutional inputs without express lawful approval. A claim of commercial confidentiality does not answer these questions. It may call for controlled disclosure, but cannot justify reliance without evidence.

Cybersecurity diligence should cover secure development, access and key management, encryption, vulnerability handling, independent testing, patches, continuity, recovery, dependencies, and incidents. Relevant threats may include data or model poisoning, prompt injection, malicious files, extraction, insecure plugins, log leakage, and compromise of a connected registry. A certification or test summary provides evidence only for its stated scope, date, exclusions, and configuration. Use abroad does not establish suitability for Pakistani records, languages, institutions, or threats.

A demonstration cannot replace repeatable evaluation. Tender documents should specify datasets or scenarios for independently controlled testing, prevent bidder access to hidden test material, and set requirements for repeatability, latency, availability, statistical uncertainty, and failure reporting. Synthetic or properly controlled test data should be used where live personal data are unnecessary. If local validation requires live or sensitive records, the pilot needs its own legal basis and controls for minimisation, security, access, retention, and deletion. Vendor claims about provenance, performance, security, and subcontractors should become warranties linked to verification, correction, suspension, and termination. Filing marketing material with the bid is insufficient.

Contractual auditability and an evidence architecture

An AI contract should specify the supplier's deliverables and the evidence available to the public institution, reviewer, and affected person. Rule 46 requires agencies to retain procurement proceedings and associated documents for at least five years. Rule 47 requires post-award publication of bid-evaluation and award documents, permitting only proprietary or public-interest information approved by PPRA to be withheld. Rule 35 separately requires an evaluation report explaining acceptance or rejection before award. These are substantial transparency duties, but they have limits. A five-year procurement file does not automatically satisfy operational-retention needs. An award record is not a model decision log, and an EPADS transaction trail does not identify the model, threshold, data, or prompt that produced an individual's output.

The contract should include a documentary assurance schedule. It should specify which records must be created, their format, owner, location, integrity controls, and retention period, together with access tiers, delivery intervals, and disposal arrangements on exit. For a consequential output, the minimum record may include time, transaction and case identifiers, input sources and relevant data versions, and the model, rules, thresholds, prompt, or configuration version. It may also include the output, meaningful confidence or uncertainty information, human review, departures and reasons, downstream action, access history, and later correction. Logs must be time-synchronised, tamper-evident, searchable, exportable, and understandable without a proprietary console. Retention must accommodate audit, limitation periods, complaints, and review without being indefinite by default. Legal holds and records duties should be reconciled with required deletion.

Audit clauses should allow the procuring agency and authorised independent testers to examine documentation, validation records, version histories, security and incident material, relevant facilities or controlled environments, and personnel. The vendor should have to produce material lawfully demanded by PPRA, the Auditor-General, a competent regulator, tribunal, or court, and help explain and reproduce relevant results. Articles 168–171 of the Constitution and sections 8 and 14 of the Auditor-General's Ordinance, 2001 support audit of legality, purpose, and expenditure, together with access to relevant public records and information. If a private vendor holds essential evidence, the institution should secure its production through the contract. The Auditor-General's direct authority over a particular vendor cannot be assumed beyond the governing law.

For public bodies covered by the federal Right of Access to Information Act 2017, the Act requires record maintenance, indexing, and proactive publication of specified material. It treats policies, public-expenditure transactions, contracts and agreements, and final orders and decisions as public records, subject to statutory exclusions and exemptions (ss. 1(2), 4–7, 16). A confidentiality clause cannot override the Act.

Contracts should classify information specifically, provide for a public version or severable schedule where feasible, and permit legally required disclosure. Legitimate trade secrets, security details, personal information, and protected intellectual property may need controlled access or severance. A general claim of vendor secrecy should not prevent explanation or review of a public decision.

Procurement remedies serve a different purpose from remedies for affected citizens. Rule 48 creates a pre-contract grievance committee and a PPRA appeal route for challenges to tender terms, evaluation, and other procurement acts. Rule 49 requires contracting parties to resolve post-contract disputes through contractual arbitration consistent with Pakistani law. Neither arrangement, without more, provides a complaint or merits-review procedure for someone denied a benefit or flagged by the system. The agency must preserve applicable administrative, sectoral, ombuds, tribunal, judicial, and information-correction routes. The contract should support these through response times, evidence provision, correction propagation, expert attendance, and implementation of lawful remedial orders. A remedy cannot depend on the vendor's commercial consent.

Change, liability, incidents, continuity, and exit

An AI service can change after award without changing its product name. The supplier may alter model weights, rules, retrieval sources, system prompts, safety filters, thresholds, data feeds, cloud regions, interfaces, or subcontractors. Rule 40 permits limited negotiation at contract finalisation without changing cost or scope. It does not allow a materially different system to be inserted after competition. Contracts should define material-change triggers and distinguish routine maintenance from changes needing notice, a documented impact assessment, repeat validation, security review, competent approval, communication with affected users, or fresh procurement. Material changes include a new purpose or population, a new data class or source, likely changes to error distribution, removal of human control, migration to another model or provider, a material vulnerability, or altered evidence access or data location. Software-as-a-service terms should prevent silent substitution where possible, require version notice and rollback, and retain the last approved configuration until reassessment is complete.

Intellectual-property terms should secure the rights needed throughout the system's life. These may include licences to use and adapt deliverables, continuing access to documentation, configurations, validation artefacts, logs, and outputs, and rights to disclose protected material to authorised reviewers under safeguards. They may also include rights to correct or transfer institutional data and, for critical bespoke systems, source-code or configuration escrow and step-in arrangements. Public ownership of every model is not always available or necessary. But a licence that expires on

termination, prohibits independent testing, or prevents the institution from understanding its own records cannot support lasting control. Contracts should address ownership of institutional inputs, annotations, fine-tuning artefacts, feedback, derived records, and generated outputs separately, and prohibit unauthorised supplier reuse.

Performance risk and liability should fall to the party able to prevent, insure against, or remedy the particular failure. Warranties can cover authority to license components, conformity with specifications, accurate disclosures, absence of undisclosed malicious code, security practices, and compliance with agreed data uses. Service credits may compensate for measurable downtime. They are unlikely to remedy unlawful disclosure, corrupted records, systemic misclassification, or obstruction of audit. Indemnities, insurance, liability caps, and exclusions should account for public loss, third-party claims, confidentiality and security breaches, intellectual-property infringement, deliberate misconduct, and data-restoration costs. The allocation will vary with the system. It cannot remove the public institution's duty to exercise lawful authority, reconsider cases, correct records, give reasons where required, or comply with courts and regulators.

An incident clause should define operational, safety, security, privacy, and rights events, including patterns of subgroup failure. It should require rapid notice, preservation of evidence, containment and rollback, root-cause analysis, correction of propagated records, notification support, regulator cooperation, and approval before resumption. Contractual notice does not replace statutory notification. The institution should be able to suspend automation and continue providing a lawful manual service.

Continuity and exit need provision at the start. Whole-life cost covers licences, data preparation, infrastructure, usage, evaluation, security, training, monitoring, appeals, transition, and decommissioning. Suppliers should maintain tested recovery arrangements, export data and metadata in documented formats, transfer configurations and necessary credentials, and assist a successor. Termination grounds may include persistent failure, material misrepresentation, unauthorised data use, a serious incident, unapproved change or subcontracting, insolvency, or inability to support audit or remedy. Public-convenience terms require separate treatment.

Offboarding requires an inventory of models, copies, embeddings, caches, logs, credentials, integrations, data feeds, documentation, and dependent decisions. It also needs a plan for open cases and appeals, lawful preservation, usable return, verified deletion when due, access revocation, downstream notice, and testing of the replacement or manual process. Delivery of software alone should not trigger an overall delivery or taking-over certificate under rule 45 for a continuing service. Acceptance, transition, final audit, and data disposition are separate milestones. Where a defect-liability or maintenance period applies, rule 45 separately provides for a defect-liability certificate.

Digital coordination leaves these responsibilities in place. Under the Digital Nation Pakistan Act, 2025, the National Digital Commission and Pakistan Digital Authority coordinate the National Digital Masterplan. PDA establishes monitoring and evaluation for Masterplan projects, reviews relevant digital plans and recommends Commission approval, and develops standards in assigned fields, including AI and data governance, without conflicting with existing regulators. It must preserve data custodians' control and responsibility (ss. 5, 8(d)–(i), 11–12). Under the National Information Technology Board Act, 2022, NITB may provide federal technical guidance, readiness review, standards, capacity support, and facilitation during conception, procurement, operation, and maintenance (s. 8(a), (c)–(d), (f), (h), (k)). These arrangements can improve specifications and shared assurance. They cannot replace the agency's competent approval, the sector institution's statutory mandate, or the accountable official's decision.

Risk-sensitive AI procurement checklist

The following checklist proposes best-practice controls for acquisition according to risk. The general PPRA framework does not expressly require every AI control listed. A control becomes binding through applicable superior law, a valid incorporated standard or direction, and/or the tender and contract. The depth of assessment should increase with consequence, scale, sensitivity, irreversibility, and connection to public authority.

Table 2.3 | Risk-sensitive AI procurement checklist

Control domain	Pre-award question and evidence	Tender or contract control	Escalation or stopping condition
Authority and purpose	Which law authorises the public function and each data use? Who approves and decides?	Purpose limitation; named controller, custodian, and decision-maker	No procurement where substantive authority is absent
Need and alternatives	What measured problem exists? Why are process reform, rules-based tools, shared services, or manual improvement insufficient?	Baseline, outcome measures, bounded pilot, review date	Stop if no measurable need or viable comparison
Population and consequence	Who is affected directly and indirectly? Which interests, locations, languages, or disabilities matter?	Use-case and excluded-use schedule; accessibility requirements	Higher approval where denial, coercion, surveillance, or safety is implicated
Performance and error	What local evidence covers false positives, false negatives, calibration, generative performance, latency, and availability?	Acceptance thresholds by context and subgroup; independent test; monitoring	Reject averages that conceal material failure; suspend on threshold breach

Control domain	Pre-award question and evidence	Tender or contract control	Escalation or stopping condition
Human control	Can an authorised, informed official review an output, depart from it, pause use, and record reasons under realistic working conditions?	Interface, staffing, training, override, escalation, and manual-continuity obligations	A nominal 'human in the loop' cannot replace effective control
Data provenance and use	Where did training, validation, fine-tuning, retrieval, and live data originate? What rights and limitations apply?	Dataset documentation; authorised-use schedule; no secondary training; quality and correction duties	Reject unexplained or unlawfully available critical data
Model provenance	Who developed and supplies each model? Which version and dependencies are used, and what limitations are known?	Model/system cards; version register; dependency inventory; warranties	Reject inability to identify or reproduce deployed version
Supplier integrity and capacity	What is known about experience, personnel, finances, incidents, litigation, debarment, beneficial ownership, and conflicts?	Accuracy warranties; continuing disclosure; key-person requirements	Disqualify materially false/incomplete qualification information under applicable rules
Subcontractors and cloud	Who can access data, alter the service, or interrupt delivery? Where is processing hosted?	Named chain; flow-down duties; prior approval for material change; data-location terms	Suspend unauthorised subcontracting or relocation
Security	What threat model, secure-development evidence, tests, patches, access controls, encryption, and recovery capability exist?	Security schedule; testing rights; vulnerability and patch deadlines; key control	Isolation or suspension on material uncontrolled vulnerability
Documentation and audit	Can output and change be reconstructed without vendor discretion?	Exportable tamper-evident logs; versions; validation and incident records; audit and regulator access	No deployment if evidence essential to review is unavailable
Transparency and reasons	What can be published, disclosed to an affected person, or reviewed under confidentiality?	Public summary; severable records; explanation support; lawful-disclosure carve-outs	Confidentiality may not oust RTI, audit, court, regulator, or remedy duties
Change control	Which updates alter purpose, population, data, performance, security, provider, or human role?	Tiered notice, approval, revalidation, rollback, and procurement-review triggers	No silent material substitution

Control domain	Pre-award question and evidence	Tender or contract control	Escalation or stopping condition
Incidents and correction	Which events and near misses must be reported? How will errors passed downstream be corrected?	Rapid notice, preservation, containment, root cause, correction, notification, resumption approval	Suspend automation while maintaining lawful service continuity
IP and portability	What rights are needed for use, scrutiny, modification, transition, and evidence preservation?	Licences, reviewer access, open exports, escrow/step-in where proportionate	Reject lock-in that prevents audit, remedy, or viable exit
Liability and remedies	Which party controls each risk, and what losses may the public or third parties suffer?	Warranties, indemnities, insurance, calibrated caps, service remedies, correction support	Do not treat service credits as exclusive remedy for serious harm
Continuity and exit	Can the institution operate through outage, supplier failure, replacement, or withdrawal?	Backups, disaster recovery, manual route, transition service, data return, verified deletion, credential revocation	No deployment without tested continuity and a funded exit plan
Lifecycle cost	Are validation, people, security, audit, appeals, updates, migration, and decommissioning budgeted?	Price schedule and cost-adjustment controls; milestone-linked acceptance	Reassess affordability where assurance or remedy is unfunded

The procurement record should allow a reviewer to answer four distinct questions. Was the acquisition competently authorised and fairly conducted? Was the system fit for its stated purpose when accepted? Did later operation and changes comply with law and contract? Can an affected person obtain evidence, correction, reconsideration, and remedy? Procurement can supply the material needed to answer each question. The label ‘AI’, a vendor's assurance, or the award of a contract cannot establish approval for every later use.

2.4 Administrative Capacity, Validation, Security, and Documentary Assurance

Software, cloud resources, and technical staff do not by themselves make an institution ready for AI. Readiness requires a lawful purpose, governed data, local validation, security, monitoring, records, reconsideration, lifecycle funding, and controlled retirement. Deployers, data custodians, users, vendors, complaint handlers, auditors, regulators, and courts all need capacity within their roles. A weakness at one point may prevent meaningful review of the system.

We must distinguish the design of an institution from its ability to operate. Pakistan's approved National Artificial Intelligence Policy 2025 proposes public-servant training, a public-sector AI register, impact assessment, audit, third-party scrutiny, redress, and human oversight for high-impact uses (Ministry of Information Technology and Telecommunication [MoITT], 2025, ss. 2.2, 3.2). These commitments establish policy direction. They do not prove that safeguards operate nationwide or that institutions have the staff, resources and procedures to carry them out. Each deployment therefore requires evidence of its own operational readiness.

Capacity as a Chain of Accountable Roles

The first requirement is a chain of accountable people with different professional skills. The programme's senior responsible officer must identify the enabling law, public purpose, affected population, consequences of error, and institution retaining final authority. Legal officers examine competence, delegation, procedure, equality, privacy, information access, records, and review. Procurement specialists convert that analysis into testable requirements, evidence rights, liability provisions, change controls, continuity, and exit arrangements. Finance officers connect expenditure with outputs, outcomes, risks, and recurring obligations. A vendor's claim that its product is 'compliant' cannot perform any of these tasks.

Data stewards and records managers have related but distinct duties. The steward should understand each source's authority, purpose, provenance, definitions, lineage, quality limits, linkages, correction and retention arrangements, and downstream users. The records manager must keep legally material versions, approvals, inputs, outputs, reasons, overrides, complaints, corrections, and changes retrievable for review. Statistical specialists design representative samples, baselines, uncertainty estimates, subgroup analysis, and monitoring. Cybersecurity staff handle threat modelling, access control, secure development, vulnerability management, incident response, recovery, and evidence preservation. Security does not establish statistical validity. Equally, an accurate model may still be used unlawfully.

Domain professionals remain necessary because statistical performance alone cannot establish clinical safety, educational validity, policing legality, eligibility accuracy, or administrative fairness. They must define meaningful outcomes, identify context missing from the data, specify when the system must abstain or escalate, and test whether it fits professional practice. Accessibility and language specialists must assess interfaces, notices, explanations, authentication, and complaint routes against the needs of people with disabilities and the affected population's language and literacy practices. Front-line users need task-specific training in intended use, limitations, uncertainty, override, escalation, documentation, and responsibility. General 'AI awareness' is insufficient. Supervisors

need the authority and time to investigate departures from outputs. Their incentives must not make an advisory output compulsory in practice.

Internal audit must remain independent of delivery while testing governance, procurement, data lineage, access, monitoring, change, and remediation. External auditors, regulators, ombuds institutions, and courts need secure access to records and appropriately independent expertise. Their roles remain distinct. An Auditor-General audit does not determine an individual's entitlement, and a court order requiring a record cannot create one that was never kept. Complaint handlers need the correct data and model version, authority to pause or reconsider, a procedure for propagating source corrections, and the means to identify recurring problems.

An institution should allocate these duties through a responsibility matrix identifying an owner, approver, consulted expert, evidence custodian, escalation route, and substitute during staff absence. Multidisciplinary review should not allow responsibility to disappear into a committee. The line institution and statutory decision-maker retain responsibility for the public function, reasons, records, and remedy, even if a vendor trains the model, a shared service hosts it, or another body supplies the data. This does not require every institution to employ every specialist permanently. Shared testing laboratories, framework contracts for independent assessment, sectoral centres of expertise, and pooled cybersecurity services can provide scarce skills. Their authority, confidentiality, independence, service levels, evidence access, and escalation duties must be clear.

Published training commitments begin the inquiry. The National AI Policy's public-sector training programme identifies a need for skills in AI, ethical use, data governance and personal-data protection (MoITT, 2025, s. 2.2). Demonstrating capacity requires evidence of filled posts, qualifications, budgets, facilities, caseloads, tests, implemented recommendations, and unresolved findings. Capacity may exist without being documented. But it has not then been demonstrated as readiness.

Total Lifecycle Cost and Public-Financial Control

The purchase price gives only part of an AI programme's cost. Before tender, a defensible estimate covers problem analysis, legal review, process mapping, data inventory and correction, digitisation, accessibility, baseline measurement, and risk assessment. Delivery adds infrastructure, connectivity, licences, integration, staff, training, validation, security testing, documentation, and parallel piloting. Operation adds monitoring, revalidation, human review, complaints, incident response, audit, vendor management, upgrades, and retraining. Retirement requires portability, migration, archives, verified data return or deletion, continuity, and decommissioning. Appeals and remedies are foreseeable costs of lawful administration. The budget must account for them.

The federal Public Finance Management Act 2019 provides important controls, although it is not an AI statute. Section 9 connects the medium-term performance budget to goals, expenditure, outputs, outcomes, indicators, and targets. Development proposals above prescribed thresholds require cost-benefit and risk assessment, independent quality assurance, technical approval, monitoring, and completion evaluation. Specified projects also require independent impact assessment (ss. 14–18). Delegation leaves the Principal Accounting Officer's (PAO) responsibility and accountability intact (s. 27). The Chief Finance and Accounts Officer (CFAO) and Chief Internal Auditor support that chain (ss. 28–29). Mid-year and year-end reports connect actual expenditure to delivered outputs and targets (ss. 34–35), and covered public entities remain subject to audit (s. 39). The Financial Management and Powers of Principal Accounting Officers Regulations 2021, amended through 27 October 2025, provide the related implementation framework (Finance Division, 2025).

These provisions can support lifecycle governance only if the AI programme is described honestly. A single capital allocation for 'automation' may hide recurring cloud charges, vendor dependence, data remediation, monitoring, complaints, and exit. Section 19 expressly addresses operation and maintenance of physical infrastructure assets. It should not be cited as a specific parliamentary rule for software maintenance. The practical point is that recurring AI obligations must appear in project design, grants, procurement documents, and medium-term estimates. The federal Medium Term Performance Based Budget FYs 2022–23 to 2024–25 illustrates a structure connecting offices, outputs, outcomes, expenditure, forward allocations, and performance indicators (Finance Division, 2022). Publication of that structure does not establish that the lifecycle cost, risk assessment, or performance indicator for a particular system is adequate.

Financial assurance must continue when the system changes. Contract variations, increased transaction volume, new datasets, additional users, model upgrades, and security responses may materially alter costs and benefits. A change request should state the fiscal effect over the remaining lifecycle, new liability and exit exposure, whether assumptions underlying competition still hold, and the appropriation and approval supporting it. Payments should depend on verified deliverables, including documentation, testing, service levels, remediation, and portability. Installation or vendor-reported accuracy alone is inadequate. Contingency funds should cover a parallel manual service, emergency suspension, incident investigation, corrective notice, and reprocessing of affected cases.

Under Articles 168–171 of the Constitution and the Auditor-General's Ordinance 2001, the Auditor-General of Pakistan (AGP) may audit covered expenditure, receipts, accounts, and bodies. AGP may require accounts, books, papers, and other documents relevant to covered transactions and call for information (ss. 8, 14). Its Financial Audit

Manual provides an institutional audit framework, but publication of a manual does not establish specialist capacity in every audit office (AGP, 2012). AGP reports go before the relevant legislature. Under rule 203, the Public Accounts Committee examines whether money was legally available and used for its authorised purpose, and whether expenditure complied with governing authority. Where expenditure exceeds a grant, the Committee examines the circumstances and makes the recommendation it considers appropriate (National Assembly of Pakistan, Public Accounts Committee, n.d., r. 203). This scrutiny is essential for value for money and failures of control. It normally reports and scrutinises rather than corrects an individual's automated decision, so internal reconsideration and external complaint or judicial routes must remain available.

Uneven Capacity and Proportionate Implementation

Capacity differs across sectors, provinces, districts, and local bodies. A central agency may have security specialists, reliable connectivity, mature records, and procurement counsel. The front-line office may work with intermittent connectivity, paper records, one overburdened operator, and no authority to correct the master database on which its decisions depend. A nationally uniform system can therefore produce unequal legality and performance in practice. Central hosting cannot supply local language assistance, domain judgement, record correction, or an accessible appeal. Nor can a national accuracy figure show whether the system fails in a province, institution, or population poorly represented in validation data.

The response should combine minimum requirements with implementation suited to local conditions. Every high-consequence deployment needs lawful authority, named responsibility, minimum security, data lineage, local validation, logging, notice, human reconsideration, incident response, audit access, continuity, and exit. Testing depth, independence, reporting frequency, and approval level should increase with consequence, scale, sensitivity, irreversibility, and concentrated dependency. An institution unable to meet the minimum should not turn an experiment into consequential use. It may use a sandbox or shadow mode, retain the existing process, limit the purpose or population, obtain a shared service, or postpone deployment until it can demonstrate the necessary capacity.

Pakistan's federal structure affects the support available. Federal or national bodies could develop interoperable technical baseline documents and provide shared services for sectors within their statutory mandates. Provincial organisations could adapt workflows, validation, professional standards, complaint mechanisms, and record correction to their substantive responsibilities. The Computer Emergency Response Team Rules 2023 set out a national, sectoral, and organisational CERT structure. Provincial CERTs would cover provincial government agencies, with local-government CERTs falling under the provincial CERT. Commencement, however, depends on

notification of an effective date in the Official Gazette. The research underlying this chapter identified no such notification by the cut-off date. Unless that notification is identified and cited, these provisions are described here as the published 2023 architecture on which National CERT practice relies, rather than unqualifiedly as operative legal duties. The design allows coordination without presuming equal capability at every level.

A shared service must identify the entity that monitors and responds to incidents, the authority over collected data and decision-making processes, the form in which provincial or local reviewers receive evidence, and who funds remedial action. Without these arrangements, sharing capacity may also make responsibility harder to see.

Kazi et al. (2020) used a mixed-methods approach to examine 51 digital-health projects in Pakistan. Reported barriers included users' difficulty with technology, costs, privacy concerns and coordination between stakeholders. Sixteen projects used AI or machine learning. These findings describe the projects and period studied, rather than national readiness for consequential public-sector AI. They support attention to implementation conditions but cannot establish the staffing, infrastructure, resources or oversight capacity of a particular institution.

Jan et al. (2024) conducted qualitative interviews with 27 senior leaders representing 27 Pakistani universities. They identified financial resources, human skills, institutional size, existing technology, and data-management practices as relevant to AI readiness. These findings concern university libraries and information services, rather than the overall administrative capacity of Pakistan's public institutions. Read together with the digital-health study, these findings support local assessment. They do not provide a national capacity assessment or establish which public AI systems are in use beyond the settings and periods studied.

Capacity assessment should use observable indicators for the particular institution. These include filled roles and staff turnover, protected operating funds, data-quality and complaint backlogs, system inventories, validation and audit reports, incident detection and closure times, exercise and recovery results, procurement variations, override patterns, accessibility testing, vendor concentration, and implementation of findings. Where no report is published, the assessment should record documentary uncertainty. A national strategy, tender, training announcement, or project entry in the Economic Survey cannot supply the missing evidence.

Validation for Pakistani Populations and Administrative Settings

Validation makes a limited claim about performance. It concerns a specified system, version, threshold, data pipeline, interface, human workflow, population, institution, and period, tested against a defined legal and administrative purpose. The body making

that claim must first identify acceptable error conditions. A false positive in fraud screening, a false negative in clinical triage, a mistranslated legal notice, and an invented answer from a public information service have different consequences. They require different measures, escalation procedures, and tolerances. A vendor benchmark or deployment abroad cannot establish fitness for Pakistani law, records, languages, infrastructure, or institutional practice.

Pre-deployment evaluation should examine five areas. Technical validation tests discrimination, calibration, robustness, uncertainty, missing data, out-of-distribution behaviour, and error rates using measures suited to the function. Data validation examines representativeness, provenance, labels, temporal coverage, leakage, and matching. Domain validation asks whether the outputs express legally and professionally meaningful categories. Human-factors validation examines comprehension, time pressure, automation bias, accessibility, authority to depart, and fallback. Security validation tests models, APIs, data flows, dependencies, access, adversarial inputs, and recovery. A team sufficiently independent of the vendor and delivery team should use version-controlled material and data held apart from development, compare performance with the administrative baseline, and retain confidence intervals, limitations, excluded groups, residual risks, and the approval decision. NIST's voluntary AI Risk Management Framework organises activities under govern, map, measure, and manage. It has legal effect in Pakistan only through a competent instrument or contract, and cannot replace local validation (Tabassi, 2023).

Population analysis needs to be planned before harm occurs. Where lawful, necessary, and statistically supportable, testing should separate relevant results by sex or gender, age, disability, province and region, urban or rural setting, socioeconomic and documentation status, and language, including meaningful intersections. National averages should not hide small groups. At the same time, unreliable estimates from small cells should not be presented as proof. Evaluation data collection must itself respect purpose, sensitivity, security, and minimisation. If the sample is inadequate, the proper finding may be ‘not validated for this group’. The response may involve restricted use, additional human review, further data collection, or a non-automated route. Equivalent performance cannot simply be assumed.

External validation needs to take place wherever the operating context differs from the testing context, including hospitals, districts, record systems, and professionally curated datasets. A shadow pilot allows generated outputs to be compared with actual decision inputs and/or outcomes while limiting exposure to potentially harmful consequences. Approval should specify use cases, user groups, sites, target populations, acceptable performance thresholds, and when administrators should abstain from using the system's guidance. It should also set conditions for returning to a manual process, performance indicators, and an expiry or review date. Once the system operates, monitoring should

track changes in data types and distributions and in the concepts those data represent. It should examine predictive performance, accuracy, calibration, subgroup disparities, complaints, overrides, response times, availability, and detected security breaches. It should also assess whether the administrative policies governing use have changed since validation.

Material change requires revalidation without waiting for the next annual review. Triggers include a new model, version, vendor, subcontractor, training or fine-tuning corpus, prompt, or retrieval source. They also include changes to source data, labels, schemas, matching methods, thresholds, outputs, interfaces, explanations, purpose, legal criteria, population, location, transaction volume, professional workflow, or human authority. Triggers also include new infrastructure or access, a serious incident, vulnerability, drift, disparity, complaint pattern, or change in governing law. Each trigger should produce a logged request, legal and risk review, proportionate independent testing, approval by a named owner, and a versioned release and rollback plan. Training, notices, contracts, and documentation should be updated. An emergency security patch may precede full testing only under a recorded emergency procedure followed promptly by retrospective review.

Security, Complaints, and a Reconstructable Assurance Record

Subject to the commencement qualification above, the published Computer Emergency Response Team Rules 2023 allocate functions across national, sectoral, provincial, local, and organisational levels. Incident tracking and ticketing with a formal audit trail are part of CERT incident-management support (r. 12(2)(a)). Rule 12 does not impose an independent logging duty on every entity. Sectoral CERTs must notify specified processes and seek compliance from functional entities. These cover information-security policy, asset classification and custody, vulnerability testing, secure development and code review, information-systems audit, continuity and incident plans, post-incident records, and a senior information-security officer (r. 22). Rule 23 addresses governance structures, funding, human resources, escalation, and an information-security management system at functional-entity level. Under rule 25(2), a functional entity must send a summary of an identified incident to its sectoral CERT and National CERT within one hour, updating it as information develops. Rule 25(3) assigns the later post-incident report to the sectoral CERT. This published architecture does not itself establish equal detection or response capacity across the public sector.

Published cybersecurity instruments need precise legal classification. The National Cyber Security Policy 2021 proposes an information-assurance framework for cybersecurity audit and compliance across public and private entities (Ministry of Information Technology and Telecommunication, 2021). This is an executive policy objective, not proof that every institution has implemented an audit regime. A

government entity or third-party auditor must identify the competent directive, relevant sector law or contract that makes an individual control applicable. Compliance still requires evidence. A cybersecurity audit cannot establish a model's legality, statistical validity, accessibility, non-discrimination or professional fitness. Equally, a fairness evaluation cannot establish secure configuration or incident-response capability. The two inquiries should inform each other while retaining their distinct methods and legal purposes.

Sector-specific regulations may impose additional duties. The NADRA Digital Identity Regulations 2025 (S.R.O. 1956(I)/2025), published in the Gazette on 14 October 2025, address accessibility, multilingual provision, security, audited access to digital identities, digital-identity audits, logging, and incidents (rr. 9–10, 11(3), 16(2)–(3), 25, 27(1)(h)-(i), 27(1)(l), 28(1)(d), 30, 32, 34–35). An institution must therefore identify the sector-specific regulatory framework governing its system. Comparison with a generic checklist is insufficient.

Public bodies covered by the federal Right of Access to Information Act 2017 must maintain adequate records and proactively publish information on their functions, governing instruments, policies and procedures, decision criteria and processes, budgets, and final performance, audit, and evaluation reports (ss. 1(2), 4-7, 16). The Act supports transparency. It cannot produce technical documentation that was never retained, correct a database, or reconsider a decision on its merits.

The lifecycle must include a workable complaint process for affected people. Notice should identify the deciding institution, explain the system's use and consequences intelligibly, state significant input sources and the adverse outcome, and offer accessible assistance. Online channels must be supplemented by assisted, offline, disability-accessible, and relevant-language routes. Each complaint needs a stable identifier, time limits, preservation of the applicable version and logs, interim protection where necessary, and reconsideration by a person with the information and authority to change the outcome. Corrections should reach the source custodian and appropriate downstream systems. People should not have to circulate between the data owner, platform operator, vendor, and agency. Monitoring must incorporate patterns in complaints, overrides, reversals, and repeated errors. Internal handling leaves competent external review available.

Documentary assurance allows an authorised reviewer to reconstruct what was proposed, authorised, funded, acquired, supplied, tested, changed, observed, decided, challenged, corrected, and retired. The institution must plan this record when the project begins, addressing retention, integrity, access controls, legal privilege, security classification, and personal-data minimisation. Keeping every log indefinitely is neither necessary nor safe. The institution should preserve legally and evidentially material information, protect it against alteration and inappropriate access, and maintain a chain

of custody for incidents and disputes. The matrix below identifies a minimum record for high-consequence systems. Sector law, the consequences involved, and limitation periods may require more.

The same evidential approach applies throughout this section. Policy targets, published controls, budget allocations, job descriptions, and audit frameworks must be dated, classified by legal status, and assessed against evidence of actual use. A proposal may identify a useful administrative arrangement without establishing a binding duty or nationwide implementation. Public-sector AI becomes reviewable when lawful authority, people, funding, validation, security, records, scrutiny, complaints, and remedies work together through identifiable administrative responsibilities.

Documentary-Assurance Record for a High-Consequence System

Table 2.4 | Documentary-Assurance Record for a High-Consequence System

Lifecycle stage	Minimum reconstructable record	Accountable custodian or sign-off	Principal review use
Authority and purpose	Enabling provisions; competence analysis; purpose, population and consequence; alternatives; risk classification; named system and decision owners	Principal Accounting Officer / competent programme authority; legal office	Legality, jurisdiction, proportionality, scope control
Need and finance	Baseline process; measurable need; options and cost-benefit analysis; full lifecycle estimate; appropriation and approvals; outputs, outcomes and indicators; contingency and exit funding	Programme owner; finance/CFAO; approving forum	Budget scrutiny, value for money, delivery of intended benefits
Procurement and supply chain	Tender, evaluation, conflicts, due diligence, vendor claims, subcontractors, model origin, licences, audit and evidence clauses, security, liability, service levels, continuity, portability, and exit	Procuring agency; procurement and contract owner	Procurement review, contract enforcement, vendor accountability
Data	Authority, purpose, source, provenance, definitions, lineage, sensitivity, licence, sharing route, quality findings, labels, matching, corrections, retention and deletion	Data custodian and steward	Data-error diagnosis, lawful reuse, correction, audit
Model and validation	Intended and prohibited uses; architecture and dependencies; training, fine-tuning, and retrieval sources; version, build, and configuration; independent test plan, datasets, metrics, subgroup results, limitations, residual risk, and approval conditions	Model owner; independent validator; domain authority	Fitness, equality, professional safety, reproducibility

Lifecycle stage	Minimum reconstructable record	Accountable custodian or sign-off	Principal review use
Deployment and human control	Sites, dates, integrations, users and permissions; training and competency; thresholds, abstention, override and escalation; manual fallback; notices; acceptance and rollback approvals	Deploying institution and responsible decision authority	Attribution, procedural fairness, human-control assessment
Operations and decisions	Time-stamped input provenance, output, uncertainty where relevant, rule, model and version, user action, override and reasons, final decision, notification, access events, and service availability	System operator; decision record custodian	Individual review, audit, incident reconstruction
Monitoring and change	Performance, calibration, drift, subgroup results, complaints, overrides, reversals, and security indicators; change request, impact reviews, approval, testing, release, rollback, and updated materials	System owner; risk/change authority; internal audit	Material-change control, continuing validity, systemic correction
Security and incidents	Asset and dependency inventory; threat model; access and security logs; VAPT and remediation; vulnerabilities; CERT tickets and notices; forensic chain; impact, recovery, and lessons; communications with affected people	Information-security officer / CERT liaison	Incident response, regulatory and forensic evidence, resilience
Complaints and remedies	Notice, complaint and assistance records; preserved case/version; interim action; human reconsideration; source and downstream correction; reasons, remedy, referral and closure; aggregate trend analysis	Complaint authority; data custodian; programme owner	Accessible redress, merits correction, pattern detection
Audit and oversight	Internal and external plans, working papers, findings, management replies, and remediation; requests and responses involving regulators, ombuds institutions, and courts; AGP report and PAC follow-up where applicable	Chief Internal Auditor; PAO; liaison for competent reviewer	Independent scrutiny, legislative accountability, legal review
Retirement	Decision and authority to retire; continuity and migration; data return/deletion verification; archives; licence and access termination; unresolved cases, liabilities and post-retirement review	Programme and records owners; finance / procurement / security sign-off	Exit assurance, preservation of rights, avoidance of orphan systems

Chapter 2 Key Arguments

AI governance begins before training and continues after software is commissioned. An administrative record's legal significance depends on authority for collection, declared purpose, source quality and context, access and reuse rules, and the ability to correct or retire it. Biometric, behavioural, and inferred information need particular care. A system may generate consequential attributes that a person never supplied and may not know

exist. Source-record, labelling, matching, interoperability, and model errors must be distinguished. Doing so identifies who could have prevented the problem, which evidence needs preservation, and what remedy can work. It also reveals when correcting the model would leave the underlying record, and the resulting harm, unchanged.

Digital identity and master registries can make authentication, payments, record matching, and service delivery faster and more reliable. They can also spread errors. A failed biometric check, inconsistent family link, outdated address, mistranscribed name, or false match may pass into benefits, banking, taxation, health, education, policing, and local administration when receiving systems treat a shared identifier or response as conclusive. Verification must remain distinct from citizenship, legal status, entitlement, risk, and final decision authority. Lawful interoperability needs a defined purpose and sharing power, named source and receiving custodians, limited fields, matching and confidence rules, transaction and access logs, security and incident controls, and duties to correct and propagate corrections. It also requires a complaint route that does not make individuals reconstruct how the state's systems connect.

Formal access does not establish practical access. Rural distance, connectivity, device and documentation costs, disability, language, migration, poverty, gendered phone access, and digital literacy affect whether someone can authenticate, understand a notice, correct a record, or seek review. An online-only correction process may deepen the problem it is meant to resolve. Cumulative exclusion develops through institutional steps: record creation, matching, a flag or denial, notice, attempted correction, referral, and remedy. At each stage, the responsible body should be identifiable and the evidence preserved. Urgent interim relief should be available where a disputed record blocks several essential services.

Procurement is a point at which constitutional and administrative requirements can be made effective. General procurement law already requires defined specifications, disclosed evaluation criteria, competition, records, and value for money. AI-specific safeguards do not follow automatically. A lawful tender should start with the public function and measurable need, then identify the affected population, consequences, baseline process, acceptable errors, human role, data and security assumptions, validation population, and conditions for stopping or withholding deployment. Due diligence must extend to datasets, model origin, subcontractors, previous testing, vulnerabilities, conflicts, and maintenance dependencies throughout the supply chain. The procuring body must test vendor claims and exercise its own judgement.

Contracts determine whether private delivery preserves public accountability. They should secure documentation, model and dataset versions, validation records, decision and access logs, security tests, incident files, explanations, and the access needed by procuring institutions, regulators, auditors, courts, and affected people. Intellectual-property and confidentiality clauses should protect legitimate interests while

permitting legality review and remedy. Material changes to purpose, data, model, population, integration, ownership, subcontractors, or threat profile require controlled review and, where appropriate, renewed validation or approval. Portability, continuity, termination assistance, justified escrow or equivalent access, data return, verified deletion, and decommissioning records help ensure that vendor withdrawal does not prevent the administration from functioning.

Purchasing a system does not create the capacity to govern it. Reviewable public-sector AI needs legal, procurement, records-management, statistical, cybersecurity, domain, accessibility, audit, finance, complaint, and litigation-support competence throughout its life. Costs extend beyond purchase prices and licence fees to data preparation, infrastructure, staffing, training, validation, monitoring, correction, appeals, incident response, vendor management, retraining, and retirement. Requirements should reflect risk and seek consistent outcomes while allowing shared services, provincial adaptation, sectoral support, and staged implementation where capacities differ. Institutions must demonstrate reliability for the relevant Pakistani languages, scripts, regions, institutions, populations, and operating conditions. Documentary assurance records what was authorised, collected, purchased, tested, changed, observed, decided, challenged, corrected, paid for, and withdrawn. Without this evidence, administrators, auditors, courts, legislatures, and affected people cannot assess whether the system performed lawfully. They are left with an assertion that it worked.

CHAPTER 03

Neural Networks, Cybersecurity, and Data Governance in Pakistan

Introduction

Neural networks change what data governance must cover. In an ordinary administrative system, we can begin with the records: why officials collected them, who may access them, how long they keep them, and how a person can correct them. These questions still matter. But they do not cover what happens when an institution selects, labels and transforms records to train a model. Training creates parameters, a new institutional artefact. Inference produces classifications, scores, predictions and information that may never have existed in a source file. An update can then change the system's behaviour while its authorised public function remains the same. Governance must follow the data into training, the model into use, and its output into the human and institutional process that gives it legal effect.

This chapter connects Chapter 2's discussion of public data infrastructure and procurement with Chapter 4's examination of consequential decisions, rights, evidence and remedies. We need enough technical understanding to ask useful legal questions. Weights, activation functions, loss, backpropagation, validation, overfitting, thresholds and generalisation reveal choices, evidence and possible failures. Convolutional, recurrent, autoencoding and transformer architectures learn different relationships. They also present different opportunities for attack and require different validation and explanation. None confers legal authority. A model may generate an alert or a score, but an identifiable institution and authorised person must decide whether it justifies restricting access, investigating someone, interrupting a service or taking another consequential action.

Cybersecurity makes the need to connect these questions clear. Neural networks may classify traffic and malware, detect unusual behaviour, analyse sequences of events, authenticate users or help decide which incidents need attention first. The security system can itself fail, however. It may misclassify, drift, leak information or suffer an attack. A false positive may exclude a legitimate user, lead to an investigation or occupy scarce analysts. A false negative may allow compromise, exfiltration or interruption of an essential service. Data poisoning, adversarial inputs, backdoors, model extraction,

inversion, supply-chain compromise and insecure updates therefore raise more than performance concerns. When institutions cannot establish which version operated, who changed it, which data shaped it or why someone acted on its output, procurement, access control, evidence preservation, incident reporting, supervision and remedy have also failed.

Pakistan's position needs to be examined instrument by instrument. Constitutional duties, statutes, delegated regulations, sectoral supervisory instruments, procurement rules, licences, contracts, policies, standards and technical frameworks have different legal force. At the manuscript cut-off, Pakistan had relevant cybercrime law, digital-government and identity legislation, sectoral cybersecurity controls, procurement machinery, CERT arrangements, and approved national policies. These did not amount to an integrated code for the neural-network lifecycle. Detailed or mandatory wording in the draft data-protection instrument does not make it enacted law either. For each duty, we must ask whether it legally applies to the institution and activity, forms part of a contract, follows an administrative direction by a competent authority, or offers comparative guidance only.

The chapter follows the system throughout its life. Section 3.1 examines data, training, model creation, validation, deployment, inference, decision, monitoring, incident, modification and decommissioning. Section 3.2 considers neural networks in security-critical administration and explains what error rates and thresholds mean for institutions. Section 3.3 connects adversarial and privacy threats with procurement, forensic evidence, responsibility, incident governance and remedy. Section 3.4 separates technical explanations from legal reasons and explains auditability through the ability to reconstruct model behaviour and the institutional decision. Section 3.5 examines how Pakistan's legal, policy, sectoral, procurement and institutional arrangements meet these requirements. Foreign and international frameworks serve as comparative resources.

The inquiry moves from records to data, models and inferences, and then to their governance throughout the lifecycle. The questions are practical. What was authorised? Which data, model, threshold and version were used? Who could act or intervene? What security and validation evidence existed, and what was logged? We also need to establish where an affected person, auditor, regulator, tribunal or court can obtain review and an effective remedy. Assigning responsibility to "the AI" answers none of this. Responsibility remains with the institutions, officials, professionals, developers, integrators, vendors, data custodians and cybersecurity providers whose actions and omissions make the system work.

3.1 Neural Networks and the Transformation of Data Governance

A neural network changes the object that the law must govern. In a conventional administrative database, a disputed assertion can usually be traced to a field, source record, custodian and amendment history. A neural network places a learned artefact between those records and action. Training turns examples into parameters, and inference applies the parameters to a new input. The resulting classification, score, prediction, generated representation or anomaly signal may never have existed in a source file. We still need to know who collected, held, disclosed, corrected or deleted a record. We must also identify who authorised the learning task, selected and labelled examples, approved the architecture and loss, validated the model, set the threshold, controlled updates and interpreted the result. Someone must remain able to correct the institutional action that follows.

There is no need to make this change mysterious. A neural network has no independent legal personality or public authority. It is a computational component with parameters, operating through data custodians, developers, vendors, infrastructure, interfaces, professional users, decision-makers, reviewers and affected persons. The difficulty is that a model can contain many choices with legal consequences, while making those choices hard to reconstruct after deployment. Calling a decision the work of "the algorithm" obscures this responsibility. Institutions need to preserve a traceable chain of responsibility from the lawful purpose through to remedy.

The minimum technical account

The basic unit, commonly called a neuron, takes numerical inputs and combines them using learned weights and a bias. In simplified form, it calculates $z = w \cdot x + b$, then applies an activation function to produce $a = \varphi(z)$. Weights determine the strength of each input's influence; the bias shifts the calculation. The activation is often nonlinear, allowing the network to represent relationships that purely linear operations, stacked together, could not. The input layer receives the chosen representation of a record, image, packet, event or text. Hidden layers transform it, and the output layer produces task-specific values, such as class scores, a numerical estimate or a reconstructed input. "Deep" usually means that the network has several learned representation layers. It says nothing about depth of legal or causal understanding (Goodfellow et al., 2016; LeCun et al., 2015).

These basic details already raise governance questions. A model sees the features designers have encoded, rather than a person, network event or administrative file in its full social and legal context. Designers may replace a missing value with a default, tokenise text, divide a log into time windows, convert an address into coordinates or represent a category numerically. Each step can remove context, introduce a proxy or hide a type of error. The output design matters just as much. Two classes may force an

uncertain case into "safe" or "unsafe". A multi-class taxonomy may group legally different events together, while a threshold may turn a continuous score into an adverse category. Reviewers therefore need the input schema, preprocessing, missing-data rules, feature construction, output definition and post-processing. A model name is insufficient.

Forward propagation passes an input through the layers, using the current parameters to produce an output. During training, a loss function evaluates that output. This mathematical objective measures poor performance on the training examples according to a chosen rule. Backpropagation efficiently calculates how parameter changes would affect the loss. An optimisation method, commonly a form of gradient descent, then adjusts weights and biases to try to reduce it (Rumelhart et al., 1986). Repeating the process across many examples produces the trained parameters.

Loss does not mean public harm. A standard loss function may treat all incorrect training examples alike, although the institution does not. It may reward overall accuracy while accepting too many wrongful denials for a small population, or improve detection while producing more false alerts than staff can manage. Class weights, sampling, regularisation, stopping rules, learning rates, batch sizes and other hyperparameters also shape the model. Every official need not reproduce the calculus. But a competent technical team must explain how the objective and evaluation metrics serve the authorised purpose, and a legally responsible owner must decide which errors are tolerable. Developers can optimise an objective given to them. They cannot decide what the enabling law, equality obligations, professional duties or acceptable administrative burdens permit.

Different learning methods place responsibility at different points. In supervised learning, the model learns associations between inputs and target labels supplied with the examples. Labels such as "malware", "eligible", "fraud" and "high risk" may look factual. Often they reflect earlier institutional decisions, incomplete investigations, measurement choices or disagreement among experts. A model may learn that labelling practice accurately without learning the underlying legal or professional concept. The record must therefore show who defined each label, how it was assigned and reviewed, which period and population it represents, and whether the label is lawful and suitable for the later use.

Unsupervised learning does not provide the same kind of target label. The system learns structures in the inputs, such as clusters, compressed representations or regularities that help identify unusual observations. Human judgment remains involved. Designers choose the features, similarity measures, architecture, training population and rule that turns distance or reconstruction error into an "anomaly". Semi-supervised learning combines a smaller labelled set with a larger unlabelled one. This can help where expert labels are costly, as with security logs, but mistakes and omissions in the labelled subset

may shape the organisation of the unlabelled material. Related self-supervised methods derive prediction targets from the data itself before adaptation to a particular task. When a public institution acquires a pretrained model from another entity, the provenance and limits of the pretraining corpus and base model remain relevant, even if it provides only a small fine-tuning dataset (Goodfellow et al., 2016; Schwartz et al., 2022).

Inference has a particular technical meaning here. It generally means applying a trained model to an input to produce an output, and should be distinguished from legal inference and the final decision. A classifier may return category scores, a prediction model may estimate a future event or continuous value, and a representation model may generate an embedding for another component. These outputs are analytic or evidentiary artefacts, not observed facts. Their admissibility and weight remain legal questions. A useful score need not be calibrated as a probability, and a high class score does not prove that the class's legal criterion is satisfied. Neural networks can be accurate yet poorly calibrated, so confidence claims require separate testing (Guo et al., 2017). The institution must define whether an output supplies information, triggers investigation, recommends, ranks, creates a presumptive result or acts automatically. It must also identify the official authorised to make the consequential determination.

Training, validation, testing, and bounded reliability

Good training performance does not prove fitness for use. The training set adjusts model parameters. Developers use a separate validation set to compare architectures, tune hyperparameters, select thresholds and decide when to stop. They reserve a genuinely held-out test set for estimating the selected system's final performance. These divisions matter. Results may look better than they are if the same people, events, near-duplicates, later observations or preprocessing information leak between sets. Repeatedly consulting the test set during design turns it into another validation set. The split must also reflect intended use. Randomly dividing records may say little about performance at a later date, in another province or institution, against a new threat family, or with a different language practice.

Overfitting happens when a model learns the peculiarities of its training material so closely that it performs poorly on new data. Regularisation, data augmentation, early stopping, dropout and other techniques can reduce the risk. None guarantees fitness (Srivastava et al., 2014). Generalisation means adequate performance on relevant unseen cases, within the limits of a task, population, environment, period and metric. A model may generalise from one benchmark yet fail on poor scans, Roman Urdu, provincial record formats, changed malware behaviour or different institutional logging practices. A procurement claim of 97 per cent “accuracy” therefore leaves much unanswered. Reviewers need the dataset, sampling frame, class prevalence, metric definition,

uncertainty, subgroup results and error costs, together with an account of how the test conditions relate to the intended administrative setting.

Validation must examine the institution as well as the technology. Technical evaluation tests discrimination, calibration, robustness, uncertainty, data leakage and behaviour outside the development distribution. Domain experts must assess whether labels and outputs reflect the relevant security, medical, educational, financial or administrative concept. Human-factors testing asks whether users understand the output and its limits, have information and authority to depart from it, and have time to investigate. Legal review examines authority for the data, criterion, consequence, procedure and decision-maker. Security validation covers the training environment, dependencies, interfaces, model artefact, inputs, outputs and update route. NIST's voluntary Artificial Intelligence Risk Management Framework treats test, evaluation, verification and validation as activities involving multiple actors throughout the lifecycle (Tabassi, 2023). This is a useful comparative assurance resource, not Pakistani law or a basis for relying on a one-time vendor certificate.

Reliability can deteriorate even when the software has no defect. Dataset shift occurs when input distributions or relationships between inputs and outcomes differ between development and use (Quiñonero-Candela et al., 2009). Attackers change tactics. Administrative eligibility rules, forms, recording practices, population behaviour and access to services also change. A feedback loop can bring earlier model outputs into later training data. Investigating flagged groups more intensively may produce more recorded positives in those groups without showing that the original score was unbiased. Operational machine-learning systems also depend on configuration, data pipelines, undeclared consumers and surrounding software. Assessing the model file alone cannot establish assurance for the system (Sculley et al., 2015).

An update can change legal consequences and evidence. Retraining, fine-tuning, or altering a label, preprocessing rule, architecture, feature, threshold, calibration method, library, retrieval source or interface may change who is affected and why. Some systems may need periodic or continual adaptation; others should remain fixed between controlled releases. The governance rule is to identify material changes, require competent approval, preserve the previous version, revalidate in proportion to risk, provide rollback and reassess notices, contracts, security and review arrangements. There is no general instruction to “always update”. The decision record must show which data, code, model and threshold governed the case when it was decided.

Architecture as a governance fact

Architecture determines how inputs interact and which dependencies a model is designed to learn. It cannot by itself establish accuracy or suitability. Four families matter particularly for security-sensitive data governance. The examples describe documented

technical uses, rather than verified deployments in Pakistani institutions (Berman et al., 2019).

Convolutional neural networks (CNNs) apply learned filters to local input regions and reuse the filters across positions. This local connectivity and sharing of weights partly explain their success in recognising images and patterns (LeCun et al., 1998). For cybersecurity, designers may construct a "spatial" representation of bytes, packet fields, traffic windows, files, images or signals. The representation has governance consequences. Rearranging fields, truncating payloads or converting binaries into images determines which local patterns the model can learn and which information it loses. An audit needs the exact encoding and preprocessing pipeline as well as the network. Validation must cover relevant malware families or traffic conditions. Security testing must also assess whether feasible or deliberately adversarial input changes can evade the classifier or change downstream behaviour.

Recurrent neural networks (RNNs) process sequences while retaining a state influenced by earlier elements. Long short-term memory networks (LSTMs) use gates and a memory state to help learn longer dependencies (Hochreiter & Schmidhuber, 1997). They may suit ordered logs, authentication events, traffic flows and attack sequences. Their usefulness depends on correct chronology, session boundaries, window length, missing events and the duration of the dependency being examined. Preserving only the latest event may make a recurrent alert impossible to reproduce. The institution also needs the ordered input window and hidden-state conditions. Explanation and audit must preserve the sequence over time: static feature-importance values may not reveal how earlier events influenced the result.

Autoencoders learn to compress an input through an encoder and reconstruct it through a decoder. This can produce useful lower-dimensional representations (Hinton & Salakhutdinov, 2006). If training mainly uses accepted "normal" behaviour, a large reconstruction error may signal an anomaly, as part of deep anomaly detection (Pang et al., 2021). But an anomaly does not prove intrusion, fraud or misconduct. It means that a particular model of normality reconstructed the input poorly. Contaminated training data, changed legitimate behaviour, or insufficient examples of a region, device, disability-related interaction or work pattern may make ordinary conduct appear anomalous. Institutions must approve and version the baseline, reconstruction metric, threshold and escalation rule, and allow them to be contested.

Transformers use attention mechanisms to build contextual relationships between elements, without depending solely on recurrence or convolution (Vaswani et al., 2017). Depending on their design and context limits, they can analyse language-based threat intelligence, event descriptions, code, and long or mixed log sequences. Governance may need to cover tokenisation, truncation, positional representation, pretraining data, fine-tuning, retrieval sources and updates to third-party base models. Attention weights

are internal computational quantities. They do not automatically explain causation or supply legally adequate reasons. The task must also be clear. Evidence that a transformer can summarise incident reports does not validate it for confidentiality classification, attack attribution or recommendations for adverse administrative action.

Architecture shapes governance in four ways. It determines the input representation and dependencies that validation must reproduce. It also affects which explanation methods and reconstruction records are useful. Different training data, sequential states, feature encodings, pretrained weights, libraries, retrieval services and update endpoints create different confidentiality and integrity risks. Finally, architecture affects procurement and exit. A public body becomes dependent if it cannot lawfully inspect or independently test the relevant artefact, reproduce a version, export necessary records or continue service after a vendor withdraws. Documentation should therefore describe the architecture and components in enough detail for an authorised reviewer. This does not mean that publishing all weights or sensitive security details is always lawful or prudent.

From record governance to data–model–inference governance

Ordinary data controls still apply: lawful collection, defined purposes, restricted access, accurate source records, authorised sharing, retention, correction and secure deletion. Neural networks add at least five matters to govern. The training corpus includes provenance, labels, exclusions, transformations and copies. The model artefact includes architecture, parameters, code, dependencies, ownership, access, integrity and version. Each inference includes the exact input snapshot, output, uncertainty or score, threshold and downstream use. The operational system includes interfaces, users, workflows, logs, fallback, monitoring and complaints. Change requires separate attention to retraining, configuration, security patches, replacement models and retirement.

These objects differ from rows in a register. Learned parameters encode statistical relationships across many examples. Deleting a source record does not automatically erase its influence on a model already trained on it. Depending on the system, retraining or a demonstrably effective unlearning process may be necessary (Vassilev et al., 2025). A model may also generate a sensitive inference, such as suspected compromise, behavioural abnormality or likely ineligibility, that the person never supplied. No source custodian can necessarily correct it by editing one field. Governance must establish who may dispute the input, review the inference, change the model or threshold, and reconsider the legal act that followed.

In Pakistan, authority must be traced to its particular source. Possessing a database, an appropriation, an executive policy or a procurement contract does not authorise every inference a public body can technically generate and use. Article 4 requires treatment in accordance with law. Depending on the function and consequence, Articles 9, 10A, 14, 19A and 25 may engage security of person, fair process, dignity and privacy of home,

access to information, and equality. The enabling statute, valid rules, allocation of business, sector law and applicable procedure must identify the competent institution and decision-maker. The approved National Artificial Intelligence Policy 2025 can guide planning and responsible adoption. As policy, it creates no new coercive power and removes no constitutional or statutory duty (Ministry of Information Technology and Telecommunication [MoITT], 2025).

Evidence and contracts require the same distinction. Sections 3–6 of the Electronic Transactions Ordinance, 2002 recognise electronic form and address accessibility, integrity, and origin or destination information where another law requires retention. They do not establish that a model output is true. Procurement can require provenance, controlled technical access, independent testing, logs, version preservation, incident support, liability, continuity and exit arrangements. It cannot transfer an official's legally entrusted discretion or legitimise an unauthorised purpose. The deploying institution must retain enough knowledge and control to explain its decisions, produce records for competent review, suspend unsafe use, correct affected cases and enforce supplier duties.

A data–model–inference lifecycle

The lifecycle is: Data → Training → Model → Validation → Deployment → Inference → Decision → Monitoring → Incident → Retraining or Modification → Decommissioning. This gives more detail than the usual "design–deploy–monitor" sequence. Stages may overlap or recur, but each marks a change in artefact, authority or responsibility. The entries identify minimum questions. Their legal answers and the available forums in Pakistan depend on the sector, territory, institution and consequence.

Table 3.1 | A data-model-inference lifecycle

Stage	Responsible actor	Authority or decision	Security duty	Documentation and audit evidence	Failure and affected interest	Review or correction route
Data	Programme owner; source custodian; data steward	Authority to collect, obtain, link, and reuse for the defined task	Integrity, access control, minimisation, secure transfer	Source, purpose, fields, provenance, quality, corrections, lineage, dataset version	Unlawful reuse, omission, mismatch, historical bias; privacy, equality, entitlement	Source correction; institutional complaint; competent regulator, ombuds body, tribunal, or court
Training	Developer or vendor; data steward; sponsoring institution	Approved objective and permission to use data and computing environment	Restricted workspace, copy control, poisoning and leakage safeguards	Dataset and label versions; preprocessing; code; loss; hyperparameters; access log	Corrupted or unrepresentative learning; confidentiality and equal treatment	Independent data/model audit; procurement enforcement; retraining decision
Model	Model developer; model owner; integrator	Approval of architecture, dependencies, ownership, and permitted uses	Artefact signing, access restriction, dependency and repository security	Architecture, parameters or controlled access, build, dependencies, model version and limitations	Poor fit for the task, compromise, undisclosed dependency; reliable administration	Technical and security review; rejection at acceptance; supplier remedy
Validation	Independent evaluator; domain expert; legal and security reviewers	Predefined acceptance criteria and deployment recommendation	Protected held-out data and reproducible test environment	Test plan, datasets, leakage controls, metrics, calibration, subgroup and robustness results, residual risk	Overfit, miscalibration, invalid benchmark; safety, equality, due process	Reject, restrict, remediate, or escalate to competent approval authority
Deployment	Deploying institution; integrator; senior responsible owner	Statutory competence, programme approval, procurement acceptance, and defined human role	Hardening, least privilege, segregation, rollback and continuity	Release, configuration, interfaces, sites, users, training, acceptance and rollback record	Misconfiguration, scope expansion, vendor lock-in; service continuity and legality	Change board; security and procurement audit; administrative supervision

Stage	Responsible actor	Authority or decision	Security duty	Documentation and audit evidence	Failure and affected interest	Review or correction route
Inference	System operator; model service; professional user	Authority to generate and use the derived output for the stated purpose	Input/output confidentiality, service authentication, tamper-evident logging	Input snapshot, time, model and rule version, output, score or uncertainty, threshold	Inaccurate or sensitive profiling; privacy, reputation, access	Input correction, reproducible re-run, inference review, interim protection
Decision	Identified statutory or professional decision-maker	Enabling power, lawful criteria, procedure, reasons, and actual authority to depart	Segregation of duties, safe fallback, controlled disclosure	Model output plus other evidence; intervention, override, rationale, final act and notice	Automation bias, unlawful denial or investigation; liberty, property, service, fairness	Human reconsideration; complaint; sector appeal; ombuds, tribunal, or judicial review
Monitoring	System owner; risk, domain, data, and cybersecurity teams	Approved monitoring purpose, indicators, thresholds, and intervention power	Detect drift, abuse, availability problems and security threats	Performance, calibration, subgroup errors, overrides, complaints, access and security alerts	Drift, feedback loop, excessive alerts, missed harm; equality and system security	Operational escalation; internal audit; suspension or revalidation
Incident	Incident lead; deploying body; applicable CERT or sector responder; vendor	Incident plan, containment authority, reporting and notification duties	Preserve, contain, eradicate, recover, and prevent recurrence	Logs, forensic copies, chain of custody, impact, notices, decisions and supplier actions	Breach, manipulation, service failure, lost evidence; confidentiality and remedy	Incident review; regulator or auditor; affected-person remedy; court where competent
Retraining or modification	Change authority; model owner; vendor; independent validator	Fresh basis for new data or purpose; material-change and release approval	Secure build, test, signing, rollback, supply-chain review	Change request; new data / code / model / threshold; comparison tests; approval and migration	Hidden functional change, regression, new bias or vulnerability; reliance interests	Revalidation; contract action; renewed notice; reject or reverse release

Stage	Responsible actor	Authority or decision	Security duty	Documentation and audit evidence	Failure and affected interest	Review or correction route
Decommissioning	Programme, records, data, security, finance, and procurement owners	Authority for retirement, retention or disposal, migration, and continuity	Revoke access, dispose of models and copies, verify return or deletion	Retirement decision, archives, unresolved cases, migration, deletion certificates, licence closure	Orphaned data, lost evidence, unsupported service; continuity and effective remedy	Records and financial audit; supplier exit enforcement; preserved case review

The matrix helps prevent three common failures in assigning responsibility. A data custodian may not make the decision, but correcting its records may be essential to remedy. A vendor may answer contractually for defective training, insecure dependencies or undisclosed changes, while the public institution remains responsible for lawful purpose, deployment, reasons and the final administrative act. Review bodies also do different work. A technical audit can diagnose the model, and internal reconsideration can correct the merits of a case. An ombuds institution or regulator may investigate within its mandate; financial audit can examine expenditure and controls; a court can review legality and require evidence. Each route has a distinct role. None replaces the others.

Neural networks extend data governance rather than remove it. Governance must cover the data–training–model–validation–deployment–inference–decision chain, including monitoring, incidents, modifications and retirement. Reliability claims apply within stated limits and to particular versions. Security reaches from source data through the model supply chain to the update channel. Correction may require both amending a record and reconsidering an inference. Institutions and people must remain able to authorise, explain, stop, repair and remedy the system. Calling a product intelligent settles little. The legal and technical question is whether every consequential step from record to action remains lawful, secure, documented, testable, attributable and open to review.

3.2 Neural Networks in Cybersecurity and Security-Critical Data Governance

Neural networks perform several distinct cybersecurity tasks. They may classify information, assess an access request, detect unusual activity, rank alerts, identify malicious files or trigger protective action. Saying that "AI provides cybersecurity" tells us little about any of them. Each task has its own evidence, errors, affected population and institutional consequences. Giving a document a provisional sensitivity label differs from blocking its transmission. Showing an analyst an anomaly score differs from suspending a citizen-facing account. Failure to authenticate a claimant does not establish that the person lacks the underlying legal status or entitlement. These distinctions must remain visible in the system's design and governance.

In Pakistan, the documentary record makes different kinds of claim, and we must keep them separate. The National Artificial Intelligence Policy 2025 proposes AI-driven threat detection and calls for human oversight in critical, high-risk operations (Ministry of Information Technology and Telecommunication [MoITT], 2025). This is approved policy direction. It does not show that a particular institution operates a neural model. The Pakistan Telecommunication Authority's (PTA) announcement of its National

Telecom Security Operations Center (nTSOC) does document a launched platform. It describes security information and event management, threat intelligence, security orchestration and automated response, and stated operator integrations (PTA, 2023). It identifies no component as a neural network and supplies no evidence of model-level performance.

Classification as a security and administrative act

Automated classification may infer that a file, message, database field, image, transaction or network flow contains personal, confidential, restricted, malicious or other security-relevant material. The model can draw on content, metadata, file structure, sender–recipient relationships, access history, device signals and behavioural patterns. Some tasks use two categories, such as malicious and benign. Others need multiple classes or labels because one record may contain identity, financial, health, security and legally privileged material together. Neural networks can help when patterns spread across many features or fixed rules cannot cover every linguistic, visual or sequential form. But the result remains an inference. The model does not create the legal category it seeks to identify.

Classification matters because a category can determine encryption, access, monitoring, retention, transmission, disclosure and disposal. Classifying material too low may expose biometric templates, investigation records, credentials or vulnerability reports. It may omit heightened logging or encryption, allow unauthorised transfer or destroy records that should be kept. Classifying too high may withhold information staff need to provide services, obstruct lawful sharing, retain unnecessary copies, impose disproportionate monitoring, delay incident response or conceal a disclosable public record. Under the federal Right of Access to Information Act 2017, the Act determines record duties and exemptions. A vendor's label does not. An automated "confidential" tag may prompt review, but cannot expand an exemption or replace the responsible officer's application of the law (Right of Access to Information Act, 2017, ss. 5–7, 16–18).

The Critical Telecom Data and Infrastructure Security Regulations 2025 (CTDISR 2025) provide a Pakistani sector-specific example. PTA made them under section 5(2)(o) of the Pakistan Telecommunication (Re-organization) Act 1996, with commencement expressed to occur on Gazette notification. They require licensees to classify assets by criticality, apply proportionate controls, identify and label sensitive information for data-loss prevention, and deploy mechanisms that can automatically block or quarantine actions risking unauthorised disclosure (regs. 15–16, 42). They also require customised monitoring use cases for malware, insider activity, exfiltration, abnormal network behaviour and breaches of personally identifiable information (reg. 39). These binding telecom requirements neither mandate neural networks nor prove their use. Where a

licensee chooses a learned classifier, it must govern that model as a component used to discharge the regulatory duty.

The National Cyber Security Policy 2021 calls for a government authentication and data-protection framework that includes data classification and appropriate protective controls (MoITT, 2021, s. 3.5). This is policy direction; it does not itself prescribe every classification rule or replace sector law. Institutions should translate applicable requirements into an approved classification scheme, define each class's legal and operational effects, record model-assigned or changed labels, and authorise someone to resolve disputed and borderline cases. Periodic review should examine whether the labels and controls still fit the information and its use. Model thresholds must not silently determine retention or disclosure schedules.

From identity signals to access and consequential denial

Risk-based access control applies an approved policy using evidence about the subject, device, resource and request. Signals may include authentication strength, device posture, network location, time, privilege, recent behaviour, failed attempts, transaction patterns and threat intelligence. A neural model can estimate how far a request departs from learned behaviour. “Behavioural authentication” may examine keystrokes, touch, gait, voice, navigation or session patterns. Anomaly-based authentication asks whether an event resembles the account's or peer group's normal activity. Both can support continuing assessment after login, but neither independently proves identity. Travel, disability, ageing, injury, shared devices, intermittent connectivity, language, assistive technology and changed working patterns can all explain legitimate departures from a baseline.

Authentication, authorisation and entitlement answer different questions. Authentication asks whether a claimant has adequately demonstrated control of an enrolled identity or authenticator. Authorisation asks whether that authenticated subject, device or process may access a specified resource under approved policy. Entitlement concerns the legal right to a benefit, licence, account service, record or other substantive outcome. NIST's zero-trust architecture usefully separates subject and device authentication from authorisation, and permits dynamic access decisions using several sources. It remains foreign technical guidance, not Pakistani law (Rose et al., 2020). NIST's current authentication guidance similarly addresses assurance that a claimant is a previously enrolled subscriber. That assurance does not prove every legal attribute associated with the person (Temoshok et al., 2025).

The National Database and Registration Authority (Digital Identity) Regulations 2025 show why the distinction matters. They provide yes/no or e-KYC authentication, require notice of successful and failed attempts, and require clear reasons for failure, such as a locked identity or biometric mismatch. An on-boarded requesting entity must secure

the service, monitor performance, arrange annual technical audit, provide backup authentication and cooperate with investigations. It remains responsible for the security and accuracy of subcontracted authentication operations. Users can access defined transaction logs. Specified incidents must reach NADRA no later than one business day after awareness or reasonable suspicion (Digital Identity Regulations, 2025, regs. 24–25, 27, 30, 35). These requirements govern authentication. A negative response may justify step-up verification or a temporary safeguard. Without the governing service law and a competent decision, it cannot extinguish citizenship, benefit eligibility, financial ownership or another legal status.

The system should distinguish alerts, recommendations and actions in its code. An alert calls for an analyst's inspection. A recommendation proposes step-up authentication, quarantine or denial. An automated action changes something: rejecting a session, locking an account, isolating a device, blocking a transfer, opening an investigation or altering public-service access. Immediate containment may be justified where delay would defeat protection, for example when quarantining a suspected malicious attachment or briefly isolating a compromised endpoint. It should follow a pre-authorised rule suited to the risk. That rule should specify scope, maximum duration, rollback, protected continuity, notice where safe, and the official who can maintain or lift the restriction. Emergency network protection must not become an unreviewable decision on the merits of a person's case.

A “human in the loop” needs time, information, competence, authority and a practical alternative. The reviewer must be able to see the event, signal sources and quality, score or class, threshold, model and policy versions, known limits and related earlier actions. They must be able to gather more evidence, depart from the output without penalty, record reasons, restore access or refer the substantive case. Thousands of unexplained alerts do not allow meaningful review. Neither does an interface offering only “approve”, or review after an irreversible denial. Where immediate blocking is necessary, prompt retrospective review and an accessible correction route form part of the control design and are not optional.

Detection across traffic, logs, files, credentials, and threat intelligence

Cybersecurity models differ mainly in what they observe and what happens next. Network-traffic classifiers examine packet or flow features, including protocol, size, direction, timing, duration and connection sequences, to distinguish applications, attacks or unusual behaviour. Intrusion-detection systems may combine known-threat signatures with learned patterns intended to detect variants or previously unseen activity. Autoencoders can compress expected traffic and use large reconstruction errors as anomaly signals. Supervised classifiers learn from labelled benign and attack examples. CNNs can identify local patterns in bytes, flow matrices or transformed representations;

recurrent networks and LSTMs represent event order; transformers can connect events across longer contexts. These uses are documented, but suitability depends on the particular system, dataset, threat, throughput and operating environment (Pawlicki et al., 2022; Shone et al., 2018).

Sequential log analysis applies this approach to operating-system, application, database, identity and cloud events. DeepLog, for example, used an LSTM to model normal sequences and flag unexpected next events (Du et al., 2017). An unusual log entry may indicate attack, but it may also reflect a configuration change, software release, authorised uncommon task or logging defect. Credential-misuse detectors can relate off-hours access, impossible travel, new devices, unusual resources, rapid downloads and privilege changes. Exfiltration monitoring may examine transfer volumes, destinations, content labels, timing and departures from a user's baseline. Threat-intelligence systems can extract report entities and relationships, cluster indicators or rank alerts. Malware models may classify static bytes, imported functions, binary-derived images or observed execution behaviour. MalConv illustrates an end-to-end classifier that takes raw executable bytes as input (Raff et al., 2018).

These outputs require evidence before they can be relied on. Production protocols, attack prevalence, encryption, devices, languages, administrators and adversaries may differ from the training benchmark. A label may reflect one detection product's judgment rather than verified ground truth. New applications and policies change benign traffic; attackers adapt deliberately. Sommer and Paxson's warning about moving machine learning from closed experiments into network intrusion detection remains relevant to governance. Evaluation must examine differences in meaning between environments, error costs, adversarial adaptation and analyst workflows (Sommer & Paxson, 2010). A benchmark score is insufficient. NIST's IDPS guidance likewise treats detection, logging, attempted prevention, reporting, configuration and maintenance as parts of an operational system (Scarfone & Mell, 2007).

False positives, false negatives, and threshold authority

A false positive labels benign activity risky. A false negative misses harmful activity. Their rates answer different questions: the false-positive rate measures the share of benign events incorrectly alerted, while the false-negative rate measures the share of attacks missed. Precision measures how many alerts are actually positive; recall measures how many positive events were detected. Accuracy can tell us very little where attacks are rare. Suppose one million events contain 0.1 per cent attacks. At 95 per cent recall and a 1 per cent false-positive rate, a detector produces about 950 true alerts and 9,990 false ones. Fewer than nine per cent are genuine. This base-rate problem can overwhelm a security team despite an apparently small false-positive rate (Axelsson, 2000). With highly

imbalanced data, precision–recall analysis is generally more useful than an aggregate receiver-operating-characteristic curve (Saito & Rehmsmeier, 2015).

The costs of these errors differ, and both can be serious. False positives may lock out legitimate officials, interrupt payments or services, prompt surveillance or investigation, create suspicion, and divert scarce analysts. Alert fatigue can then cause staff to miss real incidents. If error rates vary by disability, language, region, gender, device quality, occupation or connectivity, the control distributes inconvenience and suspicion unequally. False negatives can allow account takeover, malware execution, exfiltration, financial loss, record manipulation or disruption of essential infrastructure. They can also give false assurance. "No alert" means that the configured system did not cross its threshold. It does not establish that no breach occurred.

Choosing a threshold allocates risk and resources. Developers can demonstrate how different thresholds change sensitivity and workload, but vendors cannot decide how much interruption, investigation, missed loss or harm to rights an institution may lawfully accept. The security owner, service owner and data custodian should approve the threshold and action policy after legal, operational and equality review. Where consequences affect a person, the competent administrative or regulated decision-maker should also approve them. Assets and actions may need different thresholds. Low-confidence events may be logged, medium-confidence events may prompt step-up authentication, and high-confidence events may justify device isolation. Legal entitlement may still require independent evidence and human determination. Calibration also matters: a neural-network score of 0.9 cannot be treated as a 90 per cent event frequency without calibration testing (Guo et al., 2017).

Acceptance evidence should give the full confusion matrix for each proposed threshold, event prevalence, precision, recall, false-positive and false-negative rates, calibration, confidence intervals, throughput, latency, and failure or abstention behaviour. Tests should use held-out data representative of the operating conditions and period. Where lawful and feasible, they should cover materially affected groups and sites, compare the existing process and explain label verification. Multiple-class and multiple-label tasks require per-class results: an average can hide failure on a rare but damaging threat. Statistical evaluation should be accompanied by red-team, adversarial, drift, missing-data, encryption and integration tests. Monitoring after deployment must connect metrics with analyst workload, overrides, restored accounts, confirmed incidents, complaints, service interruptions and harms found elsewhere. NIST's AI Risk Management Framework offers useful voluntary guidance on context, measurement limits, continuing monitoring and named risk ownership (Tabassi, 2023). It is not Pakistani law.

Logging, explanation, escalation, and intervention

To review an automated security action, the institution needs more than the output. For every material alert or action, it should preserve a time-synchronised event and case identifier, source system and data provenance, and the relevant feature-pipeline, model, configuration, rule and threshold versions. The record should show the score, class, confidence or abstention; access, alert or containment action; reviewing analyst and evidence; override or continuation and reasons; notifications, referrals, downstream effects and later corrections. Aggregate monitoring should retain the population denominator, estimated prevalence, error and workload measures, drift indicators, threshold changes and known blind spots. Logs must resist undetected alteration, support export, restrict access and remain intelligible without a live vendor console. Retention should reflect legal, incident, audit and limitation needs. Indefinite logging concentrates behavioural and security-sensitive information unnecessarily.

Pakistani instruments already require parts of this record. CTDISR 2025 requires telecom licensees to keep and review audit logs, customise SIEM use cases, prevent log alteration, synchronise time and retain logging information for at least one year. Licensees must send critical and high-severity alerts to nTSOC, preserve incident evidence for at least three years, and report materially serious incidents within twenty-four hours, followed by a comprehensive report (regs. 22–25, 39–40). Procurement, configuration and records schedules must add the model, feature, dataset and threshold evidence these controls may miss. Forwarding an alert does not prove its accuracy, lawful use or subsequent correction.

The explanation must fit its audience and purpose. Analysts may need events, features, rules, a comparable baseline, confidence and suggested investigative steps. System owners need aggregate failure patterns, drift, throughput and dependence on data sources. An affected person normally needs to understand the restriction, governing rule, material facts, duration and correction route. Lawful withholding may protect detection methods or an investigation. Regulators, auditors, experts, tribunals and courts may need controlled access to sensitive records. “The AI detected anomalous behaviour” supplies no adequate reason. Source-code disclosure is neither always necessary nor usually sufficient. Institutions should protect security through summaries, redaction, confidentiality rings or independent technical inspection while preserving access to the factual and institutional basis of the action.

Escalation must reflect both cyber severity and consequences for people. Technical routes should identify the security operations centre, incident response team, asset owner, data custodian, privacy or legal officer, sector regulator or CERT, vendor and executive risk owner. When an alert blocks authentication, payment, employment access or public services, a separate administrative route is needed. Repeated low-confidence alerts about one person, rising denials, subgroup disparities, unexplained drift, missing

logs or vendor updates should trigger escalation even without a single “critical incident”. NIST's current incident-response guidance usefully places preparation, detection, response, recovery and learning within organisational risk management (Nelson et al., 2025). In Pakistan, however, the applicable instrument must determine the reporting recipient, deadline, authority and remedy.

Intervention should permit safe pause, override, rollback, isolation, alternative authentication, manual service, source-data correction, threshold adjustment, model suspension, incident declaration and referral. Authority for each action must be clear. An analyst may quarantine a file without having power to decide eligibility. A vendor may supply evidence but should not finally determine a citizen's complaint. A SOC may recommend restricting an account, while the regulated institution remains responsible for the service. Records should identify who acted, under which authority, and whether the person and downstream systems were restored. Constitutional guarantees concerning lawful treatment, fair determination of civil rights or obligations, dignity, access to information and equality apply according to context. Cybersecurity creates no independent coercive power (Constitution, 1973, arts. 4, 10A, 14, 19A, 25).

Pakistan-focused security-critical use-case matrix

The matrix sets out governance requirements for possible uses. It does not establish that Pakistan has deployed every listed system.

Table 3.2 | Pakistan-focused security-critical use-case matrix

Use case: purpose and plausible architecture	Data source; threat model; affected population	Required performance evidence	False-positive / false-negative consequence	Human role and access control	Logging and incident route	Responsible institution; vendor role; review mechanism
Sensitivity and PII classification: multilabel transformer/CNN with rules to label records and guide handling or DLP	Content, metadata, record schema and transfer context; accidental or malicious disclosure; staff, record subjects and lawful recipients	Local-language and file-type coverage; per-class precision / recall; provenance and label audit; tests on rare sensitive classes	FP: over-restriction, delayed service or disclosure, unnecessary retention. FN: disclosure, weak safeguards or wrongful disposal	Custodian approves scheme and disputed labels; least-privilege access; model cannot determine a statutory exemption	Label, score, version, rule, reviewer, transfer/block and correction; privacy/security escalation, and incident route where exposure occurs	Data-owning institution retains duty; vendor documents classifier and updates; internal review, competent information officer / regulator / auditor, and court as applicable
Risk-adaptive authentication and access: supervised or sequence model supporting step-up, session restriction or lockout	Credentials, device posture, login and behavioural signals; account takeover or insider misuse; citizens, customers, staff and contractors	Attack and legitimate-change scenarios; demographic / device / accessibility testing; calibration; fallback success and restoration time	FP: exclusion, service interruption or suspicion. FN: unauthorised access, fraud or compromise	Automatic short containment only under approved policy; trained reviewer can restore, use backup authentication, and separate identity from entitlement	Attempt, signals used, model/threshold, action, reason, notification, review and restoration; fraud/security referral	NADRA/requesting entity, bank, licensee or service body according to function; vendor supports evidence, not final status decision; complaint, sector and judicial routes preserved

Use case: purpose and plausible architecture	Data source; threat model; affected population	Required performance evidence	False-positive / false-negative consequence	Human role and access control	Logging and incident route	Responsible institution; vendor role; review mechanism
Network intrusion and anomaly detection: autoencoder, CNN, LSTM or transformer feeding IDS/SIEM	Packet/flow telemetry, asset and threat intelligence; external attack, lateral movement or novel anomaly; users and dependent services	Representative live-like traffic; prevalence; per-attack recall; false-alert volume per analyst; latency, encrypted-traffic and drift tests	FP: alert fatigue, blocked traffic and resource diversion. FN: persistence, breach, outage or corrupted records	Analysts triage; high-confidence machine-speed isolation must be bounded, reversible and reviewed	Raw/derived event, model and rule versions, alert lineage, analyst disposition, containment and chain of custody; SOC / CERT / regulator route	Network operator or public body owns protection and continuity; integrator / vendor maintains detector; internal incident review, PTA/nTSOC for covered licensees, audit and courts where engaged
Sequential logs and credential misuse: LSTM/transformer or autoencoder for unusual access and privilege sequences	Identity, application, database, cloud and privilege logs; stolen credentials, malicious insider or compromised service account; staff and account holders	Detection by scenario and role; validation after organisational change; workload, subgroup and “new user” analysis; missing-log tests	FP: investigation, stigma, suspension or intrusive monitoring. FN: privileged compromise or exfiltration	Separation of alert triage, employment / disciplinary investigation and final decision; restricted analyst access	Time-synchronise d sequence, source completeness, feature/model version, case access, investigation and outcome; insider / security / legal escalation	Employer or system owner retains disciplinary and service authority; vendor must not repurpose behaviour; internal due process, professional / sector review, tribunal or court as applicable
Malware and malicious-content classification: CNN or transformer examining bytes, behaviour, attachments or text	Files, binaries, sandbox events, email and URLs; malware, ransomware, phishing or obfuscation; senders, recipients and service users	Current and novel family tests; clean-file false-block rate; adversarial / packed-file tests; latency and sandbox containment	FP: blocked legitimate software/message and interrupted work. FN: execution, spread, data loss or outage	Quarantine may be automatic; release, destructive remediation and wider investigation require authorised review	File hash, source, model/signature versions, sandbox results, quarantine / release and propagation; malware incident route	Host institution / licensee owns action; vendor supplies signatures/model and vulnerability notice; SOC/incident team, regulator/CERT and evidentiary review

Use case: purpose and plausible architecture	Data source; threat model; affected population	Required performance evidence	False-positive / false-negative consequence	Human role and access control	Logging and incident route	Responsible institution; vendor role; review mechanism
Exfiltration and threat-intelligence prioritisation: DLP classifier with graph / transformer ranking	Transfers, endpoints, content labels, destinations and external intelligence; covert leakage or low-quality / manipulated intelligence; employees, customers, sources and counterparties	Scenario-based transfer tests; source reliability; precision at review capacity; stale-feed and lawful-sharing tests	FP: blocked lawful transfer, investigation or source exposure. FN: loss of PII, state, commercial or evidentiary material	Automated block only within defined DLP rule; custodian and security officer review release; intelligence cannot alone establish misconduct	Transfer, label, rule/model/feed version, block, reviewer and final disposition; breach, regulator and evidence-preservation route	Licensee/public body owns decision and reporting; feed/model vendors disclose provenance and changes; sector review, audit, complaint and court access under safeguards

Before calling a listed use operational, the documentary record should identify its lawful institutional owner, approved purpose and action policy, site, affected population, live data sources and integrations. It should specify accepted model, rule, threshold and configuration versions, deployment and acceptance dates, trained users, escalation contacts, continuity and complaint routes, and current vendors and subcontractors. The package should also include time-stamped production logs showing that the system ran. A tender, policy commitment, laboratory benchmark, publication, demonstration or installed licence establishes less. CyberCure, for example, documents Pakistani researchers' malware-defence research system, without establishing adoption, scale, public authority, current configuration or production performance (Shahid et al., 2024). Effectiveness requires further evidence: independently examinable local validation, monitored errors and workloads, incident and correction outcomes, material-change history, and records showing that the required controls were actually used.

Neural networks can detect cybersecurity patterns. The harder governance questions concern what an inference may do, who accepted its errors, who can stop or correct the action, and whether the record permits investigation and remedy. Secure deployment requires a traceable institutional chain. Law and policy define the objective; technical evidence informs thresholds; authorised people control consequences; logs preserve events. The responsible institution must answer for excessive alerts, missed incidents, denials, restoration and redress. Assigning the matter to “the AI” cannot discharge that responsibility.

3.3 Securing the Neural-Network Lifecycle: Adversarial Threats, Privacy, and Model Integrity

Neural-network security begins before deployment. A system learns behaviour relevant to decisions through data, software, parameters, people and infrastructure spread across a supply chain. An attacker may target training records or labels, the training process, a pretrained component, deployed weights, the prediction interface or an update made after approval. The result may resemble an ordinary mistake. A legitimate transaction appears malicious, an intrusion goes undetected, a person is wrongly associated with a risk pattern, or confidential material appears in an output. Security governance must preserve enough evidence to distinguish statistical error and drift from negligence, insider misuse, vendor failure and deliberate attack.

NIST's adversarial-machine-learning taxonomy separates attacks by lifecycle stage and by the attacker's objective, knowledge and capability. It also cautions that mitigation depends on context (Vassilev et al., 2025). For Pakistani public and regulated systems, this is an analytical resource rather than law. It helps identify who controlled the vulnerable stage, what authorised the data use or automated function, which safeguards

were promised, what evidence survives, who can suspend operation, and how affected people can obtain correction or remedy. The Prevention of Electronic Crimes Act 2016 criminalises relevant unauthorised access, copying, transmission and interference, and may address proved attacker conduct. It does not validate a compromised administrative decision retrospectively, prove the institution's due care, or provide every necessary form of notice, correction, compensation and review.

Poisoning the basis on which the model learns

Data poisoning corrupts the basis of learning. An attacker may manipulate, insert, delete, corrupt, mislabel or strategically select training material to produce a desired result. The attack may degrade overall accuracy or availability, or target a particular account, device, file pattern, locality or event class for benign or suspicious treatment. It can enter through crowdsourced samples, compromised sensors, false incident reports, contaminated threat feeds, weak record linkage, insider label changes or vendor datasets with unverifiable provenance. As the network optimises its training objective, the poisoned examples alter the weights later used for inference. The model may then run exactly as coded while applying a corrupted learned mapping.

This changes where an investigation must begin. If maliciously altered labels cause a cybersecurity classifier to miss an intrusion, the output is only the immediate result. Investigators must trace source custody, extraction, matching, labelling, preprocessing, dataset assembly and training. Reasonable diligence requires authorised-source inventories, provenance and licence records, appropriate cryptographic hashes or signatures, separated ingestion and labelling duties, and controlled write access. It also requires duplicate, outlier and label-consistency checks, representative clean holdout sets, documented sampling, and review of abrupt shifts in class balance or source composition. Joint international guidance treats data supply chains, poisoning, provenance, integrity verification, trusted infrastructure and drift as lifecycle security concerns (National Security Agency et al., 2025). No single control proves safety. With sound key binding and trust chains, a verified signature shows that a particular key holder signed an unchanged file; it cannot establish truthful content or an uncompromised signing environment. Outlier filters may remove legitimate rare events, and aggregate validation may hide targeted failures. Provenance checks therefore need behavioural tests against plausible attacker objectives.

A backdoor is a particular integrity attack. Manipulated training, fine-tuning or a pretrained model causes an otherwise normally performing system to produce the attacker's chosen output when it encounters a trigger. A visual patch, byte sequence, phrase, combination of network fields or contextual pattern may supply that trigger. Ordinary test accuracy can remain high. Conventional acceptance testing may therefore miss the compromise. BadNets showed how outsourced training and transfer learning

can introduce hidden trigger behaviour through the model supply chain (Gu et al., 2019). Later benchmarking found that reported poisoning and backdoor success depends on experimental assumptions. Vendor claims and generic benchmarks alone cannot establish security in a particular deployment (Schwarzschild et al., 2021).

Procurement must address how the attack can enter. Suppliers should disclose the origin and version of pretrained weights, datasets, libraries, annotation services, fine-tuning code and subcontractors. They should preserve reproducible build and training records and allow controlled independent testing. The deploying institution should verify artefact hashes, scan dependencies, test trigger-relevant and slice-specific behaviour, and maintain a clean rollback point. If poisoning or a backdoor affects public decisions, the authorising and deploying institution remains responsible for suspending reliance, identifying affected cases, preserving evidence and arranging reconsideration. Data providers, developers, integrators and vendors may face contractual, professional, civil, regulatory or criminal consequences according to their control, representations, fault, causation and applicable law. Describing a model as proprietary does not allocate responsibility.

Adversarial examples and evasion at inference

An adversarial example is an input deliberately designed or changed to make a model respond incorrectly. Evasion describes the attacker's objective during deployment: escape detection or secure a favourable classification without changing the trained model. Image changes may be visually slight, but cyber attacks need not be imperceptible. An attacker may alter malware bytes without changing execution, reshape traffic timing, spread activity across accounts, obfuscate a malicious string or imitate a legitimate user's behaviour. Even with strong ordinary test performance, neural networks may assign high confidence to these manipulated inputs (Goodfellow et al., 2015; Vassilev et al., 2025).

An adversarial-robustness score has limits. Its meaning depends on which features attackers can change, their query allowance, their knowledge of the architecture or gradients, and the operational constraints they must preserve. Acceptance tests should state these assumptions, include adaptive red-team work, and examine preprocessing and downstream rules alongside the classifier. Repeated boundary probing, sudden confidence changes, unusual feature combinations, differences between raw and normalised inputs, and clusters of near-threshold events can justify inquiry. They do not prove malicious intent. Investigators need the original input and acquisition context, preprocessing trace, model and threshold versions, outputs, retained confidence or scores, user and service-account logs, and associated network telemetry.

Evasion that produces a false negative may allow intrusion, exfiltration, fraud or prolonged service compromise. Excessively broad defences can instead create false positives, exclude legitimate users or intensify monitoring of unusual but lawful

behaviour. An identified risk owner must approve thresholds and fallback rules, document permitted automatic actions, and ensure that consequential lockout, investigation or denial has lawful authority and a correction route. Cybersecurity teams handle detection and containment within their mandate. The programme or sector institution remains responsible for the legal consequences of administrative action taken on the alert.

Inversion, membership inference, extraction, and leakage

Integrity attacks change model behaviour; privacy and confidentiality attacks seek information held by the model or training data. Model inversion or reconstruction uses responses, confidence values, gradients or other information to infer representative or training inputs. Attribute inference seeks an unknown sensitive attribute through other information and model access. Fredrikson et al. (2015) showed that prediction interfaces disclosing confidence information could enable reconstruction in particular settings. Membership inference asks whether a person's record appeared in training. Participation itself may be sensitive where the model concerns disease, policing, benefits or suspicious activity (Shokri et al., 2017). Such attacks need not recover an entire source file to cause disclosure. An inference interface may reveal information beyond its formal output field.

Model extraction aims to reproduce a model's functionality, decision boundary or parameters through queries or access to artefacts. For some model classes, prediction APIs can enable substitutes with high fidelity (Tramèr et al., 2016). Extraction may cause intellectual-property loss, make offline evasion research cheaper or reveal sensitive detection logic. These risks cannot make ownership and secrecy barriers to lawful scrutiny. Access should be tiered: sufficient public explanations for affected persons, fuller controlled access for the procuring body, regulator, auditor, court and qualified expert, and security safeguards against unrestricted disclosure of exploitable details.

Data leakage covers more than these attacks. Misconfiguration, compromised storage, debugging output, generated content, memorisation and malicious queries may expose training records, credentials, source code, weights, prompts, logs, cached inputs or confidential features. Researchers have recovered individual training examples from language models under demonstrated conditions. Deleting a source record therefore does not establish that a trained artefact can no longer contain or reproduce it (Carlini et al., 2021). Shared gradients in distributed learning can also permit reconstruction in some settings (Zhu et al., 2019). These findings do not mean that every model memorises or every update leaks raw data. Institutions must test the actual architecture, interface, training duplication, output controls, access patterns and threat model. Claims that data are “encoded” or “never leave the device” do not answer those questions.

Warning signs may include unusual query volumes or coverage, repeated confidence probing, reproduction of rare training strings, unexpected downloads, unusual

model-registry access, or privacy-test results exceeding approved thresholds. Investigators need authenticated API and inference logs; lawfully minimised request and response content; rate-limit events; dataset membership and lineage; training and checkpoint versions; access-control changes; exports; and privacy-test methods and results. The deployer must decide whether to withdraw and review affected outputs or decisions. The model owner and vendor must contain interface or artefact exposure, while the custodian assesses affected records and correction needs. Competent legal and incident authorities determine notification and enforcement. A contract that leaves these tasks unassigned has a governance gap.

Model compromise and the extended supply chain

Model compromise includes unauthorised acquisition or alteration of weights, configuration, thresholds, feature pipelines, orchestration code, secrets or inference infrastructure. An attacker need not defeat the learned decision boundary. Replacing the model, changing a threshold or preprocessing, stealing an API key or disabling logs may be easier. Neural networks also rely on ordinary software and infrastructure: operating systems, drivers, containers, libraries, registries, CI/CD services, cloud accounts and administrative consoles. A robust classifier with excessive privileges or an exposed registry still forms part of an insecure public system.

Supply-chain compromise can involve dataset brokers, labelling firms, pretrained-model repositories, open-source packages, managed platforms, cloud providers, cybersecurity contractors, integrators and update channels. A component may arrive malicious, be incorporated with a vulnerability, or be compromised later. NCSC secure-development guidance therefore stresses lifecycle threat modelling, supplier security, asset tracking, data and model documentation, protected infrastructure, incident processes, monitoring and update management (National Cyber Security Centre, 2023). Possible indicators include signature or hash mismatches, unexplained dependency or model-version changes, unusual privileged access, divergent builds, disabled telemetry, unapproved endpoints and newly disclosed vulnerabilities. The evidence should include asset and dependency inventories, software and model component bills where feasible, provenance attestations, signed release and deployment manifests, build logs, access records, vulnerability histories, change approvals, subcontractor maps and preserved artefacts from before and after the incident.

“Shared responsibility” needs named duties. Developers control secure design and disclosure of limits. Vendors control representations, patching, vulnerability intake and contracted evidence. Integrators control configuration, interfaces, identity and deployment; data providers control stated provenance and authorised supply; cybersecurity providers control monitoring and agreed response. Deploying institutions control lawful purpose, acceptance, operational authority, suspension, case review and

remedy. Individual users can control only matters for which they have actual information, authority, time and competence. An indemnity may redistribute financial losses between contracting parties. It cannot remove a public institution's duty to act lawfully towards an affected person.

Threat-and-accountability matrix

The matrix supports design and investigation. It does not establish that any listed attack is occurring in Pakistan. Indicators justify inquiry; attribution needs preserved evidence and competent analysis.

Table 3.3 | Threat-and-accountability matrix

Threat	Stage and attacker objective	Indicative observations	Institutional harm and essential forensic record	Principal responsibility nodes
Data poisoning	Collection, labelling, training; degrade or target learned behaviour	Provenance gaps; label conflicts; source or class shifts; hash mismatch	Wrong alerts or public action; source manifest, hashes, label history, ingestion/access logs, dataset and training version	Data provider / custodian, developer, vendor; deployer for acceptance and remedy
Backdoor	Pretraining, fine-tuning, update; trigger attacker-chosen output	Normal aggregate accuracy but trigger-specific failure; unexplained weights or source	Covert bypass or targeted flag; model provenance, checkpoints, trigger tests, build/release records	Developer/vendor and integrator; deployer for independent testing and suspension
Adversarial evasion	Inference; cross a decision boundary without changing the model	Boundary probing; crafted feature combinations; raw/preprocessed discrepancy	Missed attack or unjustified defensive block; original input, preprocessing trace, queries, outputs, thresholds	Operator / cybersecurity provider, integrator, risk owner
Inversion or membership inference	Training / deployment interface; infer attributes, examples, or participation	Confidence / gradient probing; privacy tests; systematic enumeration	Disclosure of sensitive participation or traits; API, confidence, gradient, membership and privacy-test records	Developer/vendor for interface; data custodian and deployer for lawful use and notice
Model extraction	Deployment / API / registry; clone logic or steal parameters	High-volume adaptive coverage; model download or registry anomaly	IP loss and offline attack development; query/response, identity, rate-limit, export, registry logs	Vendor/model owner, integrator, deployer security owner
Data leakage	Any stage; obtain records, secrets, prompts, weights, or memorised content	Rare-string reproduction; abnormal export; storage or output exposure	Privacy, security, evidentiary, and service harm; content, access, storage, output and version records	Custodian, developer / vendor, cloud or cyber provider, deployer

Threat	Stage and attacker objective	Indicative observations	Institutional harm and essential forensic record	Principal responsibility nodes
Model compromise	Build, registry, deployment, update; steal, replace, or alter artefacts / configuration	Signature mismatch; unauthorised privilege or threshold change; disabled logs	System-wide unreliable action; signed artefacts, CI/CD, IAM, change, deployment and telemetry logs	Integrator and infrastructure / vendor teams; deployer authorises response
Supply-chain compromise	Acquisition through maintenance; introduce malicious or vulnerable dependency	Unapproved component / subcontractor; build divergence; vulnerability notice	Hidden cross-system exposure and lock-in; component inventory, attestations, contracts, build and patch history	Prime vendor with flow-down duties; procurer / deployer for due diligence and continuity

From reasonable diligence to enforceable procurement

Reasonable diligence depends on the lifecycle stage and possible consequences. For data, it requires lawful authority, a stated purpose, provenance, integrity, sensitivity classification, restricted access, quality and poisoning checks, and correction. Training needs an isolated, auditable environment, least privilege, reproducible configuration, protected secrets, approved code and dependencies, intact checkpoints and retained validation evidence. Each model needs an owner, unique version, secure registry, documented architecture and limits, proportionate red-team and privacy testing, and controlled release. Deployment requires hardened infrastructure, authenticated interfaces, appropriate rate limits, protected telemetry, human and automated fallback, and formal acceptance of a baseline. Monitoring and change controls must detect drift and attacks, set incident thresholds and retraining triggers, allow rollback, and renew validation when data, model, purpose, population, supplier or security assumptions change materially.

Federal procurement must express these controls as requirements supported by evidence. Calling a product "secure" cannot do that work. The Public Procurement Rules 2004 govern specifications, evaluation, records and contract conditions in general, while the notified cloud-services standard document provides relevant contractual mechanisms. Neither prescribes every neural-network control. Applicable superior law, valid regulatory or administrative instruments, and precise tender and contract terms make AI-security duties enforceable against suppliers. The National Cyber Security Policy 2021 supports supplier assurance, security clauses in contracts and periodic assessment of critical suppliers (MoITT, 2021, s. 3.5). This policy direction should be translated into precise, applicable procurement and contractual obligations.

A contract proportionate to the risk should address at least these matters:

Assets, purpose and provenance: specify authorised uses, protected data and model assets, training and evaluation sources, data rights, pretrained components, licences, subcontractors and prohibited secondary uses.

Acceptance and assurance: specify performance and security tests, attacker assumptions, relevant groups and environments, consequences of false positives and negatives, independent-testing rights, remediation and rejection criteria. Certificates and benchmarks inform acceptance; they do not settle it.

Documentation and audit: require dataset and model versions, lineage, labels, configurations, training parameters, dependency and release manifests, validation and adversarial-test results, access and inference logs, retention schedules and preservation triggers. Provide controlled access for the procuring institution, competent regulators, auditors, courts and experts.

Operation and incident support: assign monitoring, vulnerability disclosure, severity classification, containment support, breach and incident-notification deadlines, forensic preservation, cooperation, root-cause analysis, correction and costs. Confidentiality terms should allow protected disclosure to legally competent reviewers.

Change and subcontracting: prohibit unapproved material changes, including data and model updates. Require notice, impact and security reassessment, revalidation, rollback and equivalent subcontractor duties. "Automatic updates" cannot automatically expand public authority.

Control, continuity and exit: settle ownership and operational control of weights, configurations, logs and derivative artefacts. Address justified portability or escrow, service continuity, termination assistance, secure return or deletion, vulnerability support and preservation of accountability records after exit.

Clauses should assign duties to those able to control the risk and produce evidence. An individual official should not answer for every hidden development defect, and suppliers should not decide whether an affected person receives a lawful remedy. Private arbitration resolves contractual disputes between parties. Administrative reconsideration, regulatory action, ombuds review and judicial relief retain their separate roles.

Federated learning: distribution is not disappearance

Federated learning allows decentralised data holders to train a shared model. A coordinator sends the current model to participating devices or institutions, which train locally. Selected updates return for aggregation, and the process repeats (McMahan et al., 2017). In Pakistan, hospitals, banks, provincial bodies or security teams could in principle collaborate across separate data holdings without pooling all raw records. This could reduce central concentration, retain some local custody and limit routine raw-data

transfers. Where the purpose and architecture are lawful, it may also support data minimisation.

The privacy benefit has limits. Raw records usually stay at participating nodes, but parameters or gradients, participation metadata, metrics and the global model move. Shared gradients have leaked training information in demonstrated settings (Zhu et al., 2019). Malicious clients can poison or backdoor the global model. Model-replacement research shows that federation does not inherently resist such attacks (Bagdasaryan et al., 2020). A compromised or malicious coordinator may manipulate distribution, participant selection or aggregation; Sybil participants may amplify attacks. Differences in local data distributions can cause unstable or unequal performance, and intermittent connectivity may systematically exclude regions or institutions. Communication overhead, endpoint patching, identity management, and clock or version inconsistencies also require security and capacity planning.

Secure aggregation can let a server calculate an aggregate without viewing each participant's update in the clear (Bonawitz et al., 2017). This reduces exposure but can make poisoning detection, debugging, audit and attribution harder. Controls should combine authenticated and authorised participation, protected channels, feasible device or workload attestation, contribution clipping, robust aggregation, minimum cohort sizes, anomaly testing and local access control. Controls should also include global validation, signed model distribution, synchronised versions and rollback. Contracts or inter-institutional agreements must assign coordinator and participant duties and specify permitted local processing, update retention, breach handling and audit access. They must identify who can exclude a node or suspend the federation.

Federation does not supply legal authority. Participants need authority to process their records for the shared purpose, coordinators need authority to aggregate and deploy, and final users need authority for consequential decisions. Removing a record or withdrawing a participant raises another difficulty: its contribution may persist in the global model. Governance must specify when future use ceases, retraining or tested unlearning where feasible, and which evidence must be retained. It must state the remaining limits candidly. Local storage alone cannot guarantee erasure.

Differential privacy: a quantified guarantee with bounded scope

Differential privacy limits how much a randomised mechanism's output distribution can change between adjacent datasets. The definition must specify adjacency and the protected unit: a record, event, user or household. In approximate (ε, δ)-differential privacy, ε bounds multiplicative privacy loss, while δ supplies additive slack or a failure parameter. Smaller values generally strengthen protection, subject to the stated unit and accounting assumptions, but may reduce utility. Differentially private stochastic gradient descent for neural networks typically clips per-example gradients and adds

calibrated noise. A privacy accountant tracks cumulative loss (Abadi et al., 2016; Dwork & Roth, 2014). Federated systems may add noise locally or after protected aggregation, each with different assumptions about trust and utility.

Differential privacy can limit membership inference and provide a testable claim stronger than informal anonymisation. It may also reduce accuracy, slow training and affect rare patterns or small groups more severely. Loss accumulates across repeated training runs or releases. A spent privacy budget cannot simply be reused. A guarantee depends on adjacency, the protection unit, clipping, sampling, accounting, correct software, randomness, side channels and who can see unnoised updates. NIST's evaluation guidance therefore treats differential privacy as a set of mathematical, implementation, data-collection, security, access-control, threat-model, utility and bias questions (Near et al., 2025). Printing one parameter in a datasheet cannot answer them.

Differential privacy does not prevent poisoning, evasion, stolen API keys, insecure storage, denial of service or theft of non-private model logic. It cannot legalise unauthorised collection, expand a permitted purpose, correct source data, provide notice or appeal, or identify the responsible official. Combining it with federated learning and secure aggregation may add protection. It also adds assumptions and may reduce forensic visibility. Public bodies should record the privacy unit, ε, δ, clipping norm, noise mechanism, accountant, release schedule, cumulative budget, validation and subgroup effects, threat assumptions and residual risks. An identified risk and data authority must approve these choices. They cannot rest solely with the vendor's data scientist.

Incident governance: from detection to reassessment

Technical response to an AI-security event must operate with institutional authority. NIST's current guidance links Detect, Respond and Recover to wider governance, identification, protection and improvement functions (Nelson et al., 2025). Neural networks require the ordinary cyber process to cover datasets, labels, model versions, thresholds, inference traces and affected decisions. The pathway below identifies a minimum documentary structure. Reporting, notification, disciplinary, procurement, sectoral and remedial duties depend on the institution and applicable law. The stages need not occur one after another. Evidence preservation begins at detection and continues alongside containment, attribution and notification.

Table 3.4 | Incident governance from detection to reassessment

Stage	Decision authority	Minimum record	Required consequence or possible remedy
Detection	Security operations under an approved incident plan	Alert source, time, affected asset/version, observable indicator, score, confidence measure or uncertainty as applicable, initial scope	Triage without treating an anomaly as proved attack; protect ongoing logs
Investigation	Named incident lead with model, data, legal, programme, and vendor support	Forensic image or protected snapshot; hashes; data/model lineage; access, query, build, training and change logs; chain of custody	Test competing hypotheses: attack, drift, defect, misuse, or record error; identify potentially affected cases
Containment	Authorised system/risk owner; emergency powers and limits recorded	Containment decision, reasons, time, commands, affected services, fallback and continuity measures	Isolate accounts, nodes, datasets, model or API; pause automated action; use a validated fallback while avoiding unnecessary service denial
Attribution	Competent investigative authority, not the model operator alone	Timeline, actor and credential evidence, supplier dependencies, uncertainty, alternative explanations	Separate technical source, legal fault, and causal contribution; refer suspected offences or contractual breach appropriately
Notification	Institution's legally competent officer with legal / CERT / regulatory input	Recipients, legal or contractual basis, content, timing, risk assessment, any delay and reasons	Notify competent authorities, suppliers and affected persons where required; give usable protective and correction information
Continuing preservation	Records custodian and legal / investigative lead	Preservation notice; immutable copies; dataset / model / configuration versions; prompts / inputs / outputs; retention and access log	Prevent routine rotation, retraining, patching, or vendor exit from destroying evidence; protect unrelated personal and security-sensitive data
Independent review	Auditor, regulator, inquiry body, court-appointed expert, or other competent reviewer	Controlled access package, methods, limitations, reproducibility record, conflicts and expert qualifications	Test root cause, diligence, vendor claims, decision impact and institutional response; permit security-sensitive closed review where lawful
Correction	Programme authority and lawful decision-maker, with technical support	Affected-case criteria, corrected records/model, notices, reconsidered decisions, reimbursements or other relief, verification of propagation	Reverse or reconsider unlawful or unreliable action; patch, retrain, revalidate, pursue contract remedies, discipline, or compensation where a legal basis exists

Stage	Decision authority	Minimum record	Required consequence or possible remedy
Reassessment	Accountable executive/risk owner with procurement and oversight input	Root-cause report, residual risk, revalidation, change approval, lessons, action owners and deadlines	Resume with conditions, maintain suspension, replace supplier, modify purpose, retrain, terminate, or decommission; report implementation and monitor recurrence

Containment can threaten evidence. A prompt patch, rollback or retraining may reduce harm while overwriting the state needed to establish what happened. Where safe, the incident plan should permit rapid capture before changes and preserve both pre-action and post-action states. If life, security or service continuity demands immediate intervention, staff should document the exception. Chain of custody must cover digital artefacts and their meaning within the model system. A hash identifies a file; a qualified witness or reproducible process explains how that file related to the deployed decision.

Fixing the vulnerability is only part of reassessment. Where poisoning affects fraud flags, missed intrusions, benefit controls, employee discipline or identity restrictions, the institution must identify the affected period, model and dataset versions, population and downstream decisions. Under the applicable authority, it should then determine whether to notify, reopen cases, restore access, correct linked records, compensate legally cognisable loss, invoke warranties or indemnities, refer professional or criminal misconduct, or decommission the system. Vendors supply evidence and repairs within their duties; they cannot determine the validity of public acts. Their involvement alone does not establish fault either. Source data, local configuration, credentials, thresholds, ignored alerts, unauthorised uses or several causes together may explain the incident. Responsibility requires reasoned analysis of the lifecycle record.

Neural-network cybersecurity therefore involves data governance, procurement, evidence, administration and remedy together. Security controls can reduce risk without taking over legal authority or responsibility. Pakistan's deploying body must be able to identify material versions, explain its confidence in suppliers, and reconstruct affected inferences and decisions. It must also be able to act during an emergency without vendor permission, preserve evidence, and provide a lawful route to correction and remedy. These capacities make the system institutionally defensible.

3.4 Explainability, Bias, Auditability, and Accountability

Calling a neural network a “black box” begins the inquiry. The phrase may describe mathematical complexity, restricted vendor access, missing institutional records or a person's inability to understand why power was exercised against them. Each problem

needs a different response. A feature-attribution chart may clarify the model while leaving procurement, data, thresholds, human use and legal authority unexplained. A readable model can still use defective labels, operate beyond its validated population or form part of an unfair procedure. We must ask whether the technical and institutional chain can be understood, tested, attributed and reviewed in light of its consequences.

In Pakistan, automated outputs operate within the constitutional and administrative order. Article 4 requires treatment in accordance with law. Article 10A protects fair trial and due process when civil rights and obligations or criminal charges are determined. Article 25 guarantees equality before law and equal protection, while Articles 14 and 19A may engage dignity, privacy of home and access to information according to context (Constitution of the Islamic Republic of Pakistan, 1973, arts. 4, 10A, 14, 19A, 25). These provisions do not establish a universal right to source code or a SHAP plot. They make authority, relevant evidence, fair procedure, intelligible justification and effective review legally significant within their respective scope and that of the applicable statute, rule, licence, professional duty or programme.

Six concepts that should not be collapsed

Technical opacity limits understanding of a model or its behaviour. Opacity may arise from scale and nonlinearity: millions of interacting parameters do not ordinarily yield a short, readable rule. Uncertainty also matters: developers may be unable to predict behaviour under distribution shift. Access restrictions create another form of opacity when weights, training data, interfaces or application programming interfaces are withheld. The difficulty depends on who is asking and what they need to know. Gradients and logs useful to a cybersecurity engineer may mean little to a benefits officer. Vendors may understand the architecture without knowing why an official adopted its output. Affected persons may need the decisive facts and rule rather than an explanation of backpropagation.

Interpretability concerns how well a particular audience can understand a model's operation or connect its inputs with its outputs. A trained user may find a short decision tree, sparse scorecard or constrained rule list inherently interpretable. Displaying parameters does not by itself achieve this. Interpretability varies with the user's knowledge, the representation and the task. Rudin (2019) argues that high-stakes systems should consider inherently interpretable models where performance permits, instead of assuming that a fallible explanation of unnecessary complexity is the only option. This is a design presumption to test. Simpler models are not automatically accurate, fair, secure or lawful.

Explainability concerns the process or artefact that answers a particular question about a model or output. It may identify influential features, approximate local behaviour, describe a counterfactual change, retrieve similar examples or summarise the

model globally. We need to assess fidelity to the system, intelligibility to the audience and relevance to the decision that audience faces. NIST's four principles similarly separate evidence or reasons, meaning for the recipient, accurate reflection of the system's process, and acknowledgement of knowledge limits (Phillips et al., 2021). An explanation for debugging, a security alert for an analyst, an audit reconstruction and a claimant's notice serve different purposes. They should not be treated as interchangeable.

Organisational transparency concerns the institution's controlled disclosure of purpose, ownership, authority, data sources, suppliers, versions, validation, thresholds, human roles, incidents, changes, limitations and complaint routes. Some information should be public or given to an affected person. Personal data, trade secrets, investigative methods and exploitable security details may justify restricting other material to authorised reviewers. Transparency does not require publishing everything. It requires arrangements that justify secrecy field by field while giving competent reviewers the evidence they need.

Auditability means being able to reconstruct and test system behaviour and the institutional decision. Source-code inspection alone cannot provide it. Reviewers need preserved versions, provenance, logs, criteria, human interventions and authority. An audit may examine whether a past output can be reproduced, validation supported the approved purpose, a vulnerability changed results, procurement claims were accurate, errors fell unequally, or officials routinely accepted recommendations without scrutiny. Different questions require different expertise and evidence. Raji et al. (2020) accordingly describe algorithmic audit as an organisational process extending from beginning to end, rather than a technical check at the last moment.

Legal accountability assigns answerability and consequences to people and institutions. It asks who held power, supplied data, designed and tested the model, set thresholds, deployed or updated it, and relied on its output. It also identifies who must reconsider a case and which forum may order or recommend relief. Possible routes include judicial review, statutory appeal, merits reconsideration, ombuds investigation, regulatory enforcement, professional discipline, public audit, contract enforcement and political oversight. Explanations and audits can support these routes without determining their jurisdiction or remedies. A model cannot hold office, give legally operative reasons, obey an order, compensate a claimant or answer for maladministration. Identifiable actors must do so.

What technical explanation methods reveal and what they do not

SHAP, LIME, attention visualisation and saliency methods can aid diagnosis and evidence when their questions, implementations and limits are documented. They cannot turn a prediction into proof of truth or legality.

SHAP. SHAP adapts cooperative game theory's Shapley values to allocate the difference between a prediction and a selected baseline among input features. In common local use, its values estimate how much each feature moved an output above or below the expected output under the method's assumptions. Aggregating local values can reveal dataset-wide patterns (Lundberg & Lee, 2017). An analyst might see that an unusual login time, a new device and a transfer pattern drove an alert. Comparing attributions across false positives may help an auditor detect reliance on a location or documentation proxy.

These values depend on choices. The background population, meaning of a “missing” feature, feature grouping, approximation and treatment of statistical dependence can change attributions materially. Correlation is particularly difficult. Credit history, address, device quality, income proxies and connectivity may convey overlapping information, allowing different assumptions to redistribute their apparent importance (Aas et al., 2021). A SHAP value does not establish causation, feature accuracy or lawful reliance. It may accurately show a model relying on a prohibited or irrational proxy. For review, institutions should preserve the implementation and version, background data, feature definitions, dependence assumptions, case input, model version, output scale and stability check. A screenshot alone leaves these questions unanswered.

LIME. LIME explains a prediction by generating changed examples around the case, querying the original model, giving greater weight to nearby examples and fitting a simpler sparse surrogate (Ribeiro et al., 2016). It can help an analyst identify a token, port or record field dominating a malware classification or risk score. Its claim remains local. The surrogate does not express the network's global rule. Fidelity depends on the definition of “nearby”, perturbation method, interpretable representation, sample number, regularisation, discretisation and random seed. Perturbed examples may also contain implausible combinations outside the real distribution. Explanations can vary between runs even when the prediction remains stable. Adversarial research has shown that perturbation-based explainers can be manipulated to hide objectionable behaviour (Slack et al., 2020). Evidentiary use therefore needs the neighbourhood, kernel, sampling and seed, surrogate fit and fidelity, software version and repeatability results.

Feature attribution and saliency. Gradients, integrated gradients, occlusion and related methods estimate which input dimensions affect an output most. Saliency visualisations highlight corresponding pixels, tokens, time steps or fields. They may help locate malicious byte patterns, unusual log intervals or spurious reliance on document borders. A plausible picture is not validation. Some methods produced similar maps after parameters or labels were randomised, showing why model- and data-randomisation “sanity checks” matter (Adebayo et al., 2018). Gradient saturation, baseline choice, preprocessing and aggregation can also change a map (Sundararajan et al., 2017). Highlighting words in an Urdu complaint or pixels near a face indicates where the

method assigned relevance. It cannot establish contextual meaning, confirm the person's identity or prove satisfaction of a legal criterion.

Attention visualisation. Attention weights show how a model component allocated attention among representations at a particular layer or head. They can reveal patterns and assist debugging, especially for sequences and language. They do not automatically explain what caused the output. In the tasks they studied, Jain and Wallace (2019) found weak correspondence with other importance measures and substantially different attention distributions yielding equivalent predictions. This does not establish that attention can never help explain a model. A heatmap describes a defined component. Treating it as an account of the whole output requires testing its fidelity for that architecture and task.

Other explanation methods have related limits. A counterfactual can state that a sufficiently different value would have changed an output while specified matters remained constant (Wachter et al., 2018). This may suggest action more clearly than attribution, yet the change may be impossible, discriminatory, causally implausible or outside the person's control. Global surrogates, feature-importance rankings, partial-dependence plots and collections of local explanations may reveal broad patterns while hiding rare or intersectional failures. Local explanations cannot prove global reliability; global averages cannot explain a particular failure. The institution should choose a method suited to its legal and operational question and test its fidelity and stability.

Explanation can create security risks. Exact fraud rules, defensive thresholds, feature weights, attack signatures and query details may assist evasion, extraction or targeting. Neighbouring records and training examples may expose personal information. But blanket claims of “security” or “commercial confidentiality” can make coercive decisions impossible to review. A proportionate arrangement should distinguish audiences and fields. Affected persons receive intelligible material facts, the output's role and a correction route; operational teams receive actionable detail; independent reviewers receive controlled technical access. Sensitive material may be redacted publicly while retained for in camera or secure examination. The sensitivity claim, withholding decision and substitute information for each recipient should themselves be recorded.

Explaining the model is not giving the legal reason

A model explanation links selected inputs to a computational output with varying fidelity. An adequate legal or administrative reason must connect lawful authority, the applicable criterion, proved or accepted material facts, evaluative judgment and outcome. It should let an affected person and reviewer understand the act and test for irrelevant considerations, factual error, improper purpose, arbitrariness and procedural unfairness. In Mustafa Impex, Karachi v Government of Pakistan (PLD 2016 SC 808),

the Supreme Court's discussion of speaking orders connected reasons with application of mind, relevant statutory material, and a rational link between facts and conclusion. The case did not concern machine learning. Its reasoning nevertheless helps explain why feature contributions alone cannot supply the administrative reason.

Suppose an application receives a fraud-risk score of 0.81, largely attributed by SHAP to address changes, failed authentication and an unusual transaction sequence. Who may deny or investigate the application? Does the governing law allow fraud risk to determine the outcome, or does it require proved ineligibility? The explanation does not establish that the address and authentication records belong to this applicant, why 0.81 triggers intervention, what contrary evidence was considered, or whether the official could and did exercise judgment. It also says nothing about correction and reconsideration. The institution's reasons must address these questions to the extent required by the governing scheme. Translating the score into fluent, model-generated prose cannot supply independent application of mind.

The necessary content and timing of reasons depend on context. Article 10A operates within its constitutional terms, while particular statutes or rules may require notice, hearing, findings or a speaking order. Article 19A and the federal Right of Access to Information Act 2017 support access and record transparency within their scope, subject to lawful exclusions and exemptions. They cannot create records never kept or replace a remedy on the merits. Security may justify limiting disclosure to an affected person, but the decision still needs a reviewable basis. At minimum, the competent decision-maker should hold the unredacted evidentiary and decisional record, and an authorised reviewer should be able to examine it.

Electronic form does not establish evidentiary weight. Sections 3–6 of the Electronic Transactions Ordinance 2002 recognise electronic form and address writing, originality and retention. Article 164 of the Qanun-e-Shahadat Order 1984 permits courts to allow evidence obtained through modern devices or techniques, while Article 59 may make expert opinion on science or art relevant. These provisions can facilitate receipt of outputs, logs and expert analysis. They do not make model outputs self-authenticating, accurate, causally explanatory or legally relevant. Reproduction may require the case input, source lineage, model and dependency versions, threshold, configuration, explanation settings, access history and integrity evidence. Experts can explain these materials. They cannot retrospectively supply a public official's missing statutory judgment.

Bias, proxies, and unequal error as lifecycle and institutional failures

Bias can enter at many stages. Historical bias arises when records reflect unequal access or enforcement. Representation bias under-represents populations or conditions. Measurement bias uses “risk”, “need”, “fraud”, “merit” or “normal behaviour” as an

inadequate proxy for the legally relevant concept. Label bias treats analyst dispositions, arrests, payment interruptions, clinical practices or earlier outcomes as neutral ground truth. Aggregation bias applies one model or threshold across groups whose data arise through different processes. Evaluation bias relies on unrepresentative benchmarks, while deployment bias assigns an output a function never validated for it (Jacobs & Wallach, 2021; Schwartz et al., 2022; Selbst et al., 2019; Suresh & Guttag, 2021).

Cybersecurity faces these problems alongside benefits, finance, policing, employment, health and education. An anomaly detector learns what counts as "normal". A baseline dominated by office networks, stable broadband, English-language traffic, newer devices or regularly documented users may flag legitimate rural, mobile, multilingual, assistive-technology, shared-device and intermittent-connectivity behaviour more often. Attackers may exploit poorly observed protocols, languages, regions or devices to avoid detection. Even a "confirmed malicious" label may reflect which alerts staff had time and authority to investigate. The model may then treat earlier resource allocation as objective evidence of threat prevalence.

Validation in Pakistan should examine language, script, province and region, rurality, disability, gender, socioeconomic position, minority status, migration and documentation gaps as possible sources of unequal performance and access. They should not become crude assumptions about social or biological identity. Comparative Urdu-model evaluation shows material variation by task and model class. This supports testing actual scripts, language practices and administrative tasks rather than relying on English benchmarks (Arif et al., 2024; Tahir et al., 2025). Biometric administration also shows why averages need exception routes. A Pakistani evaluation of biometric cash delivery identified foreseeable difficulties with hard-to-read fingerprints, including among some older people and manual labourers, as well as device and connectivity failures. It also found gains in women's control of payments and no average fall in successful collection during the studied rollout (Clark et al., 2022). Aggregate findings need subgroup evidence and accounts of how failures are handled; they do not establish uniform bias or uniform fairness.

Proxy discrimination need not involve an explicit protected attribute. Postcode, school, language, device, typing pattern, family links, documentation history, employment formality and transaction timing may correlate with region, wealth, gendered access, disability, minority identity or migration. Deleting the named field may leave that information spread across others. Detecting unequal errors may conversely require lawful collection of limited demographic data. Minimisation and security must be reconciled with evaluation, rather than make disparity impossible to measure. Testing should use necessary, protected, purpose-limited data, suitable access controls and reliable methods for small groups.

Fairness has no single metric that answers the legal question. Overall accuracy may conceal high false positives for one group or false negatives for another. Where prevalence differs, equalising one error rate may conflict with calibration or predictive value (Chouldechova, 2017). Institutions must begin with the consequences: who is investigated, locked out, delayed, denied, exposed or left unprotected; which errors can be reversed; and who can obtain human review. Relevant groups and intersections should have reported confusion-matrix measures, calibration, abstentions, missingness, uncertainty, alert-to-action rates, overrides, reversals, complaints and time to remedy. Small samples should prompt acknowledged uncertainty and possibly restricted use. They cannot justify unsupported assurance of equality.

Article 25 provides the constitutional reference for state action. Pakistan's reasonable-classification doctrine permits distinctions based on an intelligible differentia rationally connected to the object, while rejecting arbitrary classification (I. A. Sherwani v Government of Pakistan, 1991 SCMR 1041). Statistical disparity does not by itself establish every element of a constitutional violation. Equal average accuracy does not prove equal protection either. The inquiry must connect the distinction or unequal effect with the competent actor, statutory object, evidence, justification, safeguards and remedy. Articles 25(2)–(3), Article 27 for public service, disability legislation and sector-specific duties may raise separate questions. A vendor's generic "fairness score" cannot replace these legal standards.

Auditability as lifecycle reconstruction

For systems with serious consequences, an authorised reviewer must be able to reconstruct what was authorised, built, tested, deployed, observed, changed, decided, challenged and corrected. The following record identifies minimum requirements without calling for indefinite retention of every datum.

Table 3.5 | Auditability as lifecycle reconstruction

Domain	Material to preserve	Principal question enabled
Authority and purpose	Enabling provisions; approved purpose, population and consequence; intended and prohibited uses; named system and decision owners	Was the system and its role lawfully authorised?
Data and labels	Source authority and provenance; dated dataset snapshots and hashes; field definitions; matching and preprocessing; label instructions and quality; missingness, exclusions, corrections and lineage	What information and institutional judgments shaped the model?
Development and supply chain	Architecture; code/build and dependency versions; weights and configuration; training parameters and seeds where needed; developer, vendor, data-provider, integrator and subcontractor roles	Which artefact operated, and who controlled each component?

Domain	Material to preserve	Principal question enabled
Validation and approval	Training / validation / test separation; benchmark and local datasets; metrics, uncertainty, calibration and thresholds; subgroup, robustness, security and explanation tests; limitations, residual risk and signed approval	What claim of fitness was actually supported?
Deployment and human role	Release and site; permissions; interface and information shown; user training; override, escalation, abstention, fallback and rollback design	Was human control meaningful in the real workflow?
Individual output and decision	Time-stamped input snapshot and provenance; model/rule version; score, class and uncertainty; explainer, settings and output; threshold; user action and reasons; final act, notice and review route	Why did this case produce this output and legal consequence?
Monitoring, security and complaints	Drift and subgroup results; access and query logs; false-alert and missed-event investigations; incidents and forensic chain; overrides, complaints, reversals, correction and remediation	Did performance or integrity change, and was harm detected and repaired?
Change and retirement	Change request and impact assessment; new data / model / threshold; testing, approval and rollback; decommissioning, continuity, data disposition, archives and unresolved cases	Did modification or exit preserve legality, evidence and remedy?

Datasheets and model cards can organise evidence about dataset creation, intended uses, evaluation conditions, subgroup performance and limits (Gebru et al., 2021; Mitchell et al., 2019). They help document a system but do not certify it conclusively. A supplier's model card may omit the deployer's threshold, integration, workload, human practices or later updates. Independent testers should have enough access to reproduce material claims and examine foreseeable failures. Internal auditors should verify that ownership, monitoring, complaint handling and change controls work. Procurement, financial and performance audit, cybersecurity audit, rights investigation, professional review, and legality or merits review each retain their own mandate, evidentiary threshold and remedial powers.

The records themselves need protection. Integrity controls should include reliable timestamps, version identifiers, restricted access, tamper evidence, proportionate reproducible environments and forensic chain of custody after incidents. Retention should follow applicable archival, evidentiary, sectoral, contractual, complaint and limitation requirements. Blanket deletion defeats review, while indefinite logging increases privacy and security risks. Sections 3–6 of the Electronic Transactions Ordinance 2002 help establish compliance with electronic form, originality and retention requirements. The National Archives Act 1993 separately covers relevant federal public records, including machine-readable material. Neither sets a universal retention period for all model artefacts. Institutions must adopt documented schedules

that preserve evidence needed for foreseeable complaints, incidents, audits and proceedings, and securely delete material with no lawful continuing purpose.

Explanation must lead to possible correction. Affected persons need accessible notice, help in relevant languages and formats, a stable case reference, preservation of the operative version, source-record correction and human reconsideration by the legally competent body. They also need interim protection where delay would defeat relief and communication of the outcome. Systemic findings may justify threshold changes, retraining, suspension, procurement action, professional or disciplinary review, restoration of benefits or access, legally permitted compensation, or decommissioning. Public institutions cannot discharge responsibility by saying "the model decided". Vendors cannot decide whether a lawful remedy is implemented.

Technical explanations help accountable administration by showing influential inputs, local behaviour, failure patterns and security evidence. They cannot prove their own fidelity, choose the applicable legal norm, establish facts, exercise statutory discretion or allocate remedies. A named institution must connect explanation with authority, evidence, human judgment, reasons, lifecycle records, independent scrutiny and effective correction. That connection makes accountability possible.

3.5 Neural-Network Governance in Pakistan and an Integrated Lifecycle Framework

Governing the operational chain, not an abstract model

Pakistan already has law governing parts of the data–model–inference lifecycle, although no single comprehensive instrument covers it. Constitutional requirements, enabling statutes, sector regulation, procurement law and channel-specific rules may apply to different aspects of a neural-network system. Cybersecurity and AI policies set direction, technical frameworks offer controls, and contracts can preserve evidence and allocate performance risk. The difficulty is connecting instruments that bind different actors, institutions and stages. None gives "the AI" legal personality or supplies powers a public body otherwise lacks. Nor do they automatically connect a security incident with correction of the administrative decision it affected.

We must begin by identifying responsibility. A ministry or provincial department sets the public purpose; a custodian provides records; a developer chooses the architecture and training method; a vendor may supply a pretrained model. An integrator connects it to live systems, a cybersecurity provider monitors events, and a professional user interprets outputs. An authorised official then takes or approves the administrative act. Regulators, auditors, ombuds institutions, tribunals and courts may later examine different parts of this chain. Authority, evidence and review routes must remain visible

through each transfer. Governance fails when everyone can describe their limited task but no institution can reconstruct or answer for the public action that follows.

Pakistan's legal and institutional baseline

The Constitution remains the superior constraint. Depending on function and consequence, a system may engage lawful treatment, security of person, fair trial and due process in determining civil rights and obligations, dignity and privacy of home, access to information, and equality (Constitution of Pakistan, 1973, arts. 4, 9, 10A, 14, 19A, 25). These guarantees specify no neural-network architecture. They require institutions to identify lawful authority, avoid arbitrary differentiation, allow fair opportunities to contest consequential action and explain the legal decision. The Prevention of Electronic Crimes Act 2016, as amended, adds offences and investigative arrangements for unauthorised access, copying, interference, misuse of identity information and critical information infrastructure. It addresses significant attacks and misuse without supplying a generally applicable data-protection or AI-governance code (Prevention of Electronic Crimes Act, 2016).

The absence of general protection remains significant. The Ministry of Information Technology and Telecommunication's May 2023 Personal Data Protection Bill is expressly a draft. Its proposed controller duties, data-subject rights and automated-processing safeguards matter analytically, but the draft does not establish enacted obligations (Ministry of Information Technology and Telecommunication [MoITT], 2023). A draft cannot supply missing statutory authority, create a nationwide right to contest an inference or authorise coercive data use. Institutions must identify the constitutional, statutory, sectoral or contractual basis for the processing they actually undertake.

The National Artificial Intelligence Policy 2025 is approved, but remains policy rather than statute. It directs work on AI-enabled cybersecurity, human oversight in critical operations, a public-sector AI register, lifecycle impact assessments, regular audits, algorithmic accountability, redress and data security (MoITT, 2025). Budgets, implementation directions and agency performance may be assessed against these commitments. Their mandatory wording does not identify who may deny benefits, authorise surveillance, compel data transfers or decide appeals. Safeguards need implementation through enabling law, valid delegated or administrative instruments, funded processes and enforceable supplier contracts where relevant. Equally, missing implementation detail does not permit agencies to ignore constitutional or sectoral duties while waiting for AI-specific legislation.

Some enacted controls already govern particular channels. The Digital Nation Pakistan Act 2025 creates the National Digital Commission and Pakistan Digital Authority (PDA). The PDA may develop and monitor the National Digital Masterplan

and coordinate federal, provincial, local and sectoral bodies. In areas assigned by the Commission, it may set and enforce frameworks and standards for data governance, AI and emerging technology. The Act requires it to avoid overlap with existing regulators and preserve designated custodians' control and responsibility (ss. 5, 8(g)–(k), 11–12). The PDA can support common lifecycle vocabulary, assurance baselines and monitoring arrangements. A digital component does not, however, make it the health department, police authority, bank regulator or benefits decision-maker.

NADRA's regime demonstrates binding infrastructure duties with defined scope. The National Database and Registration Authority Ordinance 2000 authorises registration databases, a National Data Warehouse, interfaces, feeder and user agency links, and specified exchange and data-management functions. Feeder agencies retain updating duties (ss. 5, 7). The National Data Repository Regulations 2024 require purpose, necessity, sharing terms, security, retention and disposal for the relevant central-repository route. The National Data Exchange Layer Regulations 2025 cover federated exchange, including approval, specified services, consent with stated exceptions, minimisation, security, breach notification and logs. Originating and on-boarded entities remain responsible for their data and operations. Effective on Gazette publication on 14 October 2025, the Digital Identity Regulations 2025 govern authentication and verification notices, purpose limits, accessibility, auditable access, security assessment, logs, audit and incidents (NADRA Ordinance, 2000; NADRA National Data Repository Regulations, 2024; NADRA National Data Exchange Layer Regulations, 2025; NADRA Digital Identity Regulations, 2025). These instruments can govern model inputs and exchanges. Authentication still does not determine citizenship, benefit eligibility, culpability or risk. An exchange interface confers no substantive decision-making power.

Cybersecurity instruments add another layer. The approved National Cyber Security Policy 2021 sets direction on national risk, critical infrastructure, assurance, procurement, incident response and capacity building. It includes a Cyber Governance Policy Committee and a proposed national, sectoral and organisational structure (MoITT, 2021). It is policy. The published Computer Emergency Response Team Rules 2023, made under PECA sections 49 and 51, describe national, sectoral, provincial, local and organisational CERT arrangements, functional-entity governance, vulnerability and security processes, incident reporting, ticketing, post-incident records and audit trails. But rule 1(3) requires a further Official Gazette notification appointing commencement. The official sources reviewed to the cut-off did not identify one. This is documentary uncertainty, not proof that no notification exists. Until it is located, the Rules should be described as the published 2023 architecture on which National CERT practice relies, without unqualified claims that their duties have commenced (Computer Emergency Response Team Rules, 2023, rr. 1(3), 12, 22–25).

The National Cyber Security Policy 2021 also proposes an information-security assurance framework, testing and accreditation facilities, and security assessment of products and services (MoITT, 2021, s. 3.6). These are policy objectives. They do not establish the operation of a particular audit scheme or certify a deployed neural model. Any resulting control must have an identifiable institutional owner and an applicable legal, administrative or contractual basis. Publication of a policy establishes its content, while implementation requires evidence of budgets, staff, procedures, tests, findings and remediation.

Sector rules can impose clearer duties. PTA's current register lists the Critical Telecom Data and Infrastructure Security Regulations 2025 and marks the 2020 regulations repealed. Issued under the Pakistan Telecommunication (Re-organization) Act 1996, CTDISR-2025 requires covered licensees to assign governance responsibility, assess risk, manage acquisition and suppliers, preserve audit and incident evidence, control change, test security and report incidents (Pakistan Telecommunication Authority [PTA], 2025). These are delegated telecom regulations with a defined sectoral reach. A licensee's neural intrusion detector operates within that security regime. The same PTA mandate cannot be assumed for a provincial hospital or land registry.

The State Bank of Pakistan also acts within its financial mandate. The Enterprise Technology Governance and Risk Management Framework for Financial Institutions requires technology governance and integration of technology risk with enterprise risk management. The 2023 digital-banking measures add specified customer and channel protections, while the 2025 Technology Risk Management Framework for Payment Institutions addresses technology and cybersecurity risk for covered payment institutions (State Bank of Pakistan [SBP], 2017, 2023, 2025b). These instruments may govern a regulated entity's neural fraud or anomaly detector without naming AI. Their application depends on the regulated entity, activity and scope of each instrument. A governance framework does not by itself establish the validation, security or fair operation of a particular neural model.

Procurement supplies further controls. The Public Procurement Regulatory Authority Ordinance 2002 and current federal Public Procurement Rules 2004 govern federal acquisition and record preservation. Provincial regimes need separate examination. The notified National Standard Bidding Document for Procurement of Cloud Services and approved Pakistan Cloud First Policy 2022 address service levels, security, audit, incidents, data handling and contract management (Public Procurement Regulatory Authority [PPRA], 2024a; MoITT, 2022). They neither certify AI assurance nor authorise substantive public decisions. Contracts must secure model and dataset versions, validation and adversarial tests, logs, subcontractor information, vulnerabilities, update histories, portability, deletion, exit and assistance to regulators and courts. Procuring institutions also need staff able to specify and enforce these duties. A contract

alone does not establish equivalent assurance capacity across provinces, sectors and local bodies.

The same care is needed with claims about government databases. Legal instruments establish NADRA's National Data Warehouse and exchange infrastructure. BISP publicly describes the National Socio-Economic Registry as a household repository for programme targeting and NADRA cross-validation (Benazir Income Support Programme [BISP], n.d.). This establishes infrastructure and stated administrative uses. It does not demonstrate a neural network in a particular workflow, its performance, or whether a promotional reference to "AI" means prediction, matching or ordinary rules. Claims of operational use should identify the production owner, approved purpose, architecture or decision logic, go-live version, affected population, validation, logs and review route.

General oversight can obtain some evidence but cannot recreate records never preserved. The Right of Access to Information Act 2017 supports record maintenance and access for federal public bodies within its exclusions and exemptions. The Auditor-General's constitutional and statutory mandate covers relevant expenditure, accounts, transactions and documents (Right of Access to Information Act, 2017, ss. 4–7, 16; Auditor-General's Ordinance, 2001, ss. 8, 14). Neither can replace missing model-version logs, validation files, reasons or contractual rights of vendor access.

The table distinguishes the status and limits of these instruments.

Table 3.6 | Status and limits of governance instruments

Instrument group	Status at cut-off	Contribution	Legal boundary
Constitution and statutes	Superior or enacted law	Authority, rights, mandates, review	Controls follow function and consequence
Digital Nation and NADRA instruments	Statute and channel-specific rules	Coordination, custody, identity, exchange	No transfer of sector decision power
PECA; PTA and SBP instruments	Statute and sector regulation	Offences, risk, audit, incidents	Offence- or sector-limited
NCSP, AI and Cloud policies	Approved policies	Direction and policy objectives	No new coercive power; trace binding pathway
CERT Rules 2023	Published; commencement notification not identified	Tiered incident architecture	Commencement remains uncertain
Personal Data Protection Bill 2023	Draft	Proposed data duties and rights	Not present law

Comparative frameworks as tools of translation

Comparative instruments are useful when their safeguards can be connected to a Pakistani legal route. GDPR article 22 offers a comparator for significant decisions based solely on automated processing, including safeguards for human intervention and contestation. Article 32 links security with risk, resilience, recovery and regular testing (Regulation (EU) 2016/679, arts. 22, 32). For high-risk systems, the EU AI Act combines risk management, data governance, documentation, record-keeping, deployer information, human oversight, accuracy, robustness, cybersecurity, post-market monitoring and incident duties. It distinguishes providers from deployers (Regulation (EU) 2024/1689, arts. 9–15, 16–27, 72–73). Neither governs a Pakistani ministry merely because its vendor works internationally. Subject to actual extraterritorial application on each instrument's terms, their value is in identifying missing safeguards and questions a contract should answer.

NIST's AI Risk Management Framework uses Govern, Map, Measure and Manage to connect context, testing, responsibility and response. Its Cybersecurity Framework 2.0 uses Govern, Identify, Protect, Detect, Respond and Recover to link model security with enterprise governance (Tabassi, 2023; Pascoe et al., 2024). ISO/IEC 27001 supports information-security management; ISO/IEC 23894 provides AI-risk guidance; ISO/IEC 42001 specifies an AI management system (International Organization for Standardization [ISO], 2022, 2023a, 2023b). These voluntary or contractual resources may inform specifications, regulatory guidance, expert evidence and assessments of reasonable diligence. Certification must be read within its scope. An information-security certificate does not prove lawful training data, equal errors, intelligible reasons or a valid administrative decision.

The OECD AI Principles add accountability according to role and traceability across the lifecycle. UNESCO's Recommendation connects impact assessment and governance with human rights, diversity, inclusion and remedy (Organisation for Economic Co-operation and Development [OECD], 2024b; UNESCO, 2021). These are normative resources, not Pakistani delegated legislation. Adoption should proceed in workable parts and reflect institutional capacity. A short, enforceable tender schedule for evidence may achieve more than a broad alignment commitment that no official has the training or funding to verify.

Responsibility across the lifecycle

Responsibility should reflect legal mandates and technical control. Data providers and statutory custodians answer for authority, provenance, definitions, quality and source-record correction. Developers answer for architecture, training methods, label construction, known limits, reproducibility and security testing. Product vendors answer for representations, components, vulnerabilities, subcontractors, maintenance and

version history. Integrators answer for interfaces, configuration, identity and access controls, local-environment testing and preserved dependencies. Cybersecurity providers may detect and investigate events. That role does not allow them to determine the legality or merits of benefit denials, employment sanctions or police action.

The public function stays with the deploying institution. Its competent authority must approve the purpose, data use, threshold, acceptable residual risk, human role, suspension criteria, budget and remedy. Professional users need sufficient information, time and authority to check outputs, depart from them, record intervention and escalate uncertainty. The named official exercising statutory power must make the legal decision and give reasons. Sector regulators supervise within their mandate; CERTs coordinate incidents; auditors examine specified evidence; courts and tribunals apply their respective review standards. Contracts may allocate commercial loss and indemnity between institution and supplier. They cannot transfer constitutional or statutory duties owed to the public. A model can hold no office and bear no blame.

Before deployment, a responsibility-and-evidence schedule should name each control's owner, approver, operator, evidence custodian, escalation recipient and substitute. Joint committees can assist consultation, but individual authority must remain identifiable. If a national platform provides a model to a provincial service, the platform body may own hosting and model integrity. The provincial custodian remains responsible for its data, while the provincial service authority retains responsibility for decisions, notice, correction and remedy.

An integrated lifecycle framework

Before deployment, sponsors should identify the lawful public or regulated purpose, statutory competence, affected population and consequences. They should map data sources and inferences, including authority for reuse, provenance, labels, representativeness, correction routes and sensitivity. Threat modelling must cover poisoning, evasion, extraction, inversion, leakage, backdoors, compromised updates and supplier dependencies. Validation must remain independent of training and reflect the actual Pakistani language, region, population, institution and operating conditions. Published Urdu evaluations show why English or vendor benchmarks cannot establish local performance (Arif et al., 2024; Tahir et al., 2025).

Approval should bring together legal, privacy, equality, security and model-impact assessments, procurement diligence and documented consideration of alternatives. It should establish subgroup baselines, consequences of false positives and negatives, reviewers' competence, accessibility, complaint handling and an independent test plan. Contracts should secure evidence, control subcontracting and updates, define incident deadlines and notification assistance, preserve audit and forensic rights, allocate remediation costs, guarantee rollback and continuity, and settle exit and deletion. The

approving official should sign off the specific model, dataset, configuration, threshold and use. A generic licence for "AI" does not define that approval.

During operation, identity and access controls, least privilege, separation of development, testing and production, integrity checks and input/output protection must accompany administrative safeguards. Each material output should be linked to its model, dataset, configuration and threshold versions, input provenance, meaningful confidence or uncertainty, the professional user's action and final legal reason. Monitoring should examine drift, attack indicators, subgroup and language performance, overrides, complaints, downstream exclusion and service effects. Someone must review the dashboard, have power to intervene and record the response. Affected persons need intelligible notice and correction or review suited to the underlying decision. Feature-importance explanations from a vendor cannot supply the whole process.

After an incident, institutions should detect, investigate, contain, provisionally attribute, give legally required notification, preserve evidence, arrange independent review, correct and reassess. Incident commanders may isolate systems or revoke access. Competent programme authorities must decide whether consequential decisions should be paused, reprocessed or manually reviewed; custodians must freeze and correct compromised records; legal and procurement officers must preserve claims and invoke supplier duties. Evidence should cover dataset and model hashes, lineage, code, configuration, access and query logs, alerts, prompts or inputs, outputs, human interventions, network and authentication records, communications, timelines, affected decisions and chain of custody. Reporting to a CERT does not replace separately required notice to a regulator or affected person. Closing a security incident does not resolve an erroneous public decision. Findings should prompt proportionate suspension, retraining, supplier cure, contract remedies, downstream correction, compensation where law provides, or decommissioning.

Material updates require governance review. New data sources, label definitions, architectures, pretrained components, weights, thresholds, integrations, populations, purposes, hosting arrangements, subcontractors or security profiles can change risk and legal authority. The change owner must identify what differs; custodians and legal authorities must recheck data and purpose. Developers and independent testers must run regression, security, bias and local-validity tests, after which the institution must approve a versioned release with rollback. High-consequence systems should prohibit silent vendor updates. Where drift triggers retraining, contested outputs must not become unchecked labels that reproduce the same feedback loop.

Termination must reduce exposure while preserving service and accountability. Exit plans should migrate the public function, revoke credentials, close interfaces, inventory copies, verify vendor and subcontractor return or deletion, securely dispose of weights

and sensitive artefacts where required, and record residual dependencies. Retention must distinguish source records, training copies, model artefacts, operational logs, incident evidence and final administrative records under their respective laws. Immediate destruction may frustrate an appeal or investigation; indefinite retention increases breach and secondary-use risks. Decommissioning is complete when downstream services no longer depend on the model, unresolved complaints have a forum, affected outputs remain reconstructable for the lawful retention period, and an accountable official signs closure.

Table 3.7 | Lifecycle governance phases

Phase	Accountable decision	Principal actors	Minimum record	Trigger
Pre-deployment	Authorise use and residual risk	Sponsor, authority, custodian, supplier, legal / procurement	Authority, lineage, version, validation, threat/impact assessments, contract	Absent authority, failed validation or inadequate evidence
Operation	Continue, constrain or suspend	Institution, user, security, custodian, regulator	Versioned outputs, access, intervention, reasons, performance, complaints	Drift, unequal error, attack or nominal review
Incident	Contain, notify, preserve, remedy	Incident lead, programme authority, CERT/regulator, investigator	Images, hashes, logs, chronology, affected decisions, custody, correction	Material compromise or affected rights/services
Change	Reauthorise release	Change owner, developer, tester, legal/security approvers	Changes, regression / adversarial tests, validation, approval, rollback	New purpose, population, supplier or unresolved flaw
Termination	Preserve service and evidence	Programme / records owners, supplier, security, complaints team	Migration, revocation, disposition, return/deletion, archive, closure	Live dependency, pending case or retention conflict

From scattered controls to lifecycle accountability

Pakistan's response should fit the identified gap. Missing general rights concerning personal data, consequential inferences and automated processing require legislation specifying application, regulator powers, exceptions, correction, contestation and remedies. Standards and Cabinet frameworks cannot supply that authority. Sector gaps, such as using neural risk scores to deny access, initiate investigations or alter entitlement, may need rules on thresholds, human authority, notice, reasons, audit access and appeal made within the competent sector.

Weak procurement may need model-specific clauses and schedules rather than another regulator. PPRA and provincial procurement bodies can prepare templates proportionate to risk, while agencies remain responsible for their uses. Implementation needs funded multidisciplinary teams, shared testing facilities, incident exercises, regulator access to secure testing environments and staged deployment. National requirements should set a defensible minimum for high-consequence systems while allowing sectoral and provincial adaptation. Coordination needs referral protocols among PDA, National and sectoral CERTs, custodians, regulators, auditors and complaint bodies. Common incident and case identifiers can support this without displacing legal mandates.

Pakistan's record contains both established controls and significant gaps. It has an enacted digital-coordination authority, extensive identity and exchange rules, cybercrime and procurement law, sectoral cybersecurity regulation, an operating National CERT contact, an approved information-security framework and documented financial-sector exercises. General data protection remains incomplete. The CERT Rules' commencement is unresolved, emerging institutions are still advertising specialist posts, and public evidence of model-specific assurance across government is limited. Missing inventories, validation reports or incident records should be identified as documentary uncertainty. Their absence does not establish that a system is absent, or that it is safe.

Governance must extend from records to data, models and inferences, and from initial approval to monitoring, incidents and modifications. It must also connect technical components with the institutions responsible for them. Cybersecurity, privacy, procurement, equality, evidence and remedy meet within this lifecycle. Existing constitutional, sectoral, procurement, digital-identity, CERT and audit arrangements can accommodate many necessary duties in Pakistan. Remaining deficiencies should be named precisely: missing authority or sector rules, weak contracts, inadequate evidence access, fragmented coordination or limited capacity. This diagnosis gives Chapters 4 and 5 a concrete starting point. General commitments to “responsible AI”, or assigning responsibility to the model, cannot identify who must act.

Chapter 3 Key Arguments

First, neural networks expand what data governance must cover. Training selects and transforms data, labels embody judgments, optimisation fixes relationships in parameters, and inference produces classifications, scores, predictions or generated content that may not exist in source records. Governance must cover datasets, labels, code, weights, configurations, thresholds, validation evidence, inferences, human interventions, updates, incidents and retirement. The legally relevant unit is the connected data–model–inference–decision chain. Examining the model alone leaves consequential choices outside view.

Second, technical design determines what the model can learn, where it can fail and what reviewers can prove. Architecture suited to local patterns is not automatically suited to event sequences. An anomaly score does not establish intrusion, and strong aggregate accuracy can hide poor calibration or unequal errors among affected people. Training, validation and testing must remain distinct. Performance claims need evidence from the intended Pakistani languages, records, regions, institutions, populations, threats and operational constraints. Material changes in data, model, threshold, integration, purpose, population or threat environment require controlled authorisation, versioning, revalidation, security assessment and documentation.

Third, cybersecurity involves using neural models for protection and protecting those models. Classification, behavioural authentication, traffic analysis, intrusion detection, malware analysis and alert prioritisation can improve defensive capacity. They also introduce attack and failure points in data, training, models, inputs, outputs, interfaces, supply chains and updates. Poisoning can corrupt learning; evasion can suppress valid alerts; backdoors can cause targeted misclassification. Inversion and leakage may reveal training information, extraction may enable replication or attack, and compromised dependencies can defeat controls before deployment. These vulnerabilities become institutional and legal failures when actors cannot reconstruct provenance, access, versions, tests, logs and decision authority.

Fourth, errors have consequences beyond a performance report. Function, threshold, population and attached action determine their effects. False positives may cause lockout, surveillance, investigation, reputational harm or diversion of security resources. False negatives may allow breaches, fraud, exfiltration, unsafe access or public-service interruption. Institutions must identify whose risk they are optimising, approve thresholds, provide trained human review, preserve overrides and offer accessible correction. A nominal reviewer cannot repair automated harm without time, authority, information and the practical ability to depart from the system.

Fifth, explanation, transparency, auditability, legal reasons and accountability remain distinct. SHAP, LIME, saliency, feature attribution and attention visualisation can indicate associations and local influences. They may be unstable, sensitive to correlation, incomplete, non-causal or revealing of security-sensitive information. A feature-contribution chart cannot establish lawful authority, accurate source records, a valid policy, meaningful official consideration or justification for the final act. Review also needs data and model versions, provenance, labels, training configuration, validation and security tests, thresholds, access and event logs, changes, interventions, complaints and the official's own rationale.

Pakistan should govern the neural-network lifecycle as a whole, connecting cybersecurity, privacy, procurement, automated decisions, evidence, equality and liability. Constitutional rules, statutes, delegated instruments, regulators, procurement

controls, CERT structures, sectoral cybersecurity duties, policies and technical frameworks already supply parts of the structure. Their status, scope, commencement and institutional reach must be stated accurately. EU, NIST, ISO, OECD and UNESCO materials can inform specifications, assurance and gap analysis without becoming Pakistani law through citation. Reform should identify missing authority, incomplete sector rules, weak contracts, fragmented responsibility, inaccessible evidence and limited capacity before creating another institution. At deployment, during operation, after incidents, through material changes and at termination, an identifiable body must remain able and obliged to authorise, secure, explain, investigate, correct, compensate where law permits and decommission the system.

CHAPTER 04

Automated Decision-Making, Fundamental Rights, Legal Responsibility, and Judicial Control

Introduction

The preceding chapters examined who may govern an artificial intelligence system, how data and procurement shape its operation, and why control must continue from data collection through model use and inference. We now come to the person affected by that process. A match may stop a benefit payment. An anomaly score may bring closer scrutiny, a biometric comparison may redirect an investigation, and a recommendation may influence clinical judgement. Even the order of applications can determine who receives attention first. Technical sophistication answers only part of the legal question. We also need to know which institution acted, whether it had authority, and whether its use of information or inference was fair, supported by evidence, open to review and capable of correction.

We must begin with the documents. Pakistan has extensive digital systems for identity, finance, telecommunications, revenue, security, public services and platforms. A policy, research prototype, tender, press report or vendor description does not by itself establish that a system is in production use. This chapter uses specific terms for that reason. Announced refers to a competent institution's statement of intention or launch without evidence of continuing use. Proposed refers to a documented design or solicitation. Procured requires an award, executed contract or equivalent commitment, while piloted means a bounded trial in a defined setting or population. Operational requires credible evidence of production use. Suspended and discontinued require evidence that use has paused or ended. Unverified means that available documents cannot substantiate a material claim. Each status belongs to a particular version, institution, place, function and time. One project cannot establish national deployment.

Harm can arise at several stages. A model may misclassify a person, rank cases poorly, use an inappropriate proxy, drift or produce unequal error rates. The same adverse result may also follow from a duplicate record, an old address, a mistaken identity link or a missing field in a local language. Rigid rules, inaccessible correction procedures, weak

procurement and an official's unquestioning reliance may contribute as well. Calling every digital error an AI failure obscures the cause. Legal analysis must identify where the failure occurred and who can correct it. That may be a data custodian, programme authority, professional user, regulated provider, developer, integrator, vendor, cybersecurity provider, regulator, reviewer or court.

The rights inquiry follows what the system does. Dignity, security of person, fair trial and due process, privacy of home, access to information, equality and other constitutional or statutory protections do not depend on an AI label. Their application turns on the purpose, intrusiveness, consequences and unequal effects of the act concerned. Machine-generated profiles require particular care. An inference can become a fact on which an institution acts even though the person never supplied it and it appeared in no original record. Connected systems can then spread the injury. A single record or classification may pass through authentication, benefits, banking, licensing, policing, health, education and employment, excluding a person repeatedly.

Procedural fairness makes it possible to bring a technical output before a legal reviewer. To challenge an adverse act, a person needs notice of the decision, its legal basis and material facts, the role of automation, the responsible institution, and the route and deadline for correction or appeal. Reconsideration also requires an authorised reviewer with the records, competence, time and practical freedom to depart from the output. Different forums need different evidence. Courts, tribunals, regulators, ombuds institutions, auditors, commissions and professional bodies have distinct mandates and standards of review. An explanation of the model may help. It cannot replace source records, versions, thresholds, validation, security and incident evidence, human interventions, contractual access rights or the official's own reasons.

Responsibility depends on legal duty and practical control. The location of the source code does not settle it. Buying a vendor's system does not remove a public institution's statutory power or constitutional responsibility. Equally, the institution's continuing responsibility does not necessarily release developers, suppliers, data providers, integrators, employers, platforms or professionals from duties that otherwise apply. Those duties may arise in contract, regulation, professional practice, consumer or labour law, or civil and criminal law. We have to trace the process from data and procurement through model operation and professional reliance to the final act. The recognised harm must then be connected to a duty, the required evidence, a competent forum, the applicable standard of review and an effective remedy.

Section 4.1 examines public administration and regulated sectors, separating documented deployment from aspiration and uncertainty. Section 4.2 considers how Pakistani constitutional and statutory protections apply to surveillance, profiling, inference, discrimination and cumulative exclusion. Section 4.3 explains the notice, reasons, disclosure, preservation, authentication and professional safeguards needed to

challenge decisions and evidence influenced by algorithms. Section 4.4 traces responsibility through institutions and suppliers, identifies the available forums, and considers individual and systemic remedies in light of urgency, reversibility, scale and recurrence. Throughout, the inquiry connects rights and harm to duty, evidence, forum and remedy. Technical complexity and divided institutional roles must not leave public or regulated power without an identifiable legal actor.

4.1 Consequential Uses in Public Administration and Regulated Sectors

Establishing system status in Pakistan

Before mapping automated decision-making in Pakistan, we need evidence of what is actually happening. A speech may announce an AI-enabled programme, a tender may never result in an award, and a portal may automate only data entry. Matching or ranking may also occur without an AI label. The relevant questions concern the system's function and status. Promotional language cannot answer them.

Announced means that a competent institution has stated an intention or launch, without evidence of completed procurement, a bounded trial or continuing production use. Proposed covers a documented design, enabling provision, research plan or solicitation whose use has not been shown. An award, contract or equivalent commitment is needed for procured; a solicitation alone is labelled 'proposed tender issued'. Piloted means a trial limited by time, site or population. Operational requires credible evidence of production use, although this establishes neither national coverage nor satisfactory performance, and does not show that every advertised function works. Suspended means a temporary stop, and discontinued means use has ended. Unverified applies where the public record does not establish the claimed status or function. It does not mean that the system does not exist.

The sources reviewed did not establish a system-wide suspension or discontinuance. The Federal Board of Revenue (FBR) suspended agent licences as an enforcement measure. It did not thereby suspend Faceless Customs Assessment (FCA).

Confidence in the source is a separate question. High confidence ordinarily requires an enacted instrument or current official record, preferably supported by an audit, evaluation or observable service. A specific implementation claim or bounded study may support moderate confidence. Announcements, tenders and descriptions that omit the owner, version, geography or outcome carry low confidence. These distinctions matter. A police launch may be documented while the model, watchlist, errors and contribution to an arrest remain unverified. A digital land registry may operate without any AI valuation function.

For each use, we distinguish the source record, linkage, model or rule, output, professional assessment, final act and review. An old survey, wrong identity link or mistranscribed land record is different from a model error. Discretionary overrides, weak contracts, vendor outages and inaccessible complaint offices create further points of failure. Several may contribute to the same denial. Identifying them determines both the correction needed and the forum able to provide it.

Social protection, identity and access to public services

The Benazir Income Support Programme (BISP) provides the clearest example of an administrative process with serious consequences for its users. Its National Socio-Economic Registry (NSER) records household composition and socio-economic circumstances through computer-assisted surveys and supports targeting for cash-transfer schemes under BISP's statutory programme. BISP reports that the 2010 exercise covered 27 million households and the 2019–2021 update covered 35 million. It also acknowledges that a static registry becomes outdated, producing errors of inclusion and exclusion. A dynamic-registry arrangement with the National Database and Registration Authority (NADRA) provided for 647 registration offices where applicants and beneficiaries could report changes in household circumstances (BISP, n.d.-c, n.d.-d). The current 8171 portal allows eligibility queries by identity-card number. Partner banks release Kafaalat payments after biometric verification (BISP, n.d.-a, n.d.-b).

BISP records and World Bank programme documents support an operational classification, with high confidence, for the federal targeting, eligibility-query and biometric-payment process (World Bank, 2024). It affects applicant and beneficiary households, particularly women receiving Kafaalat payments. The programme has national reach, but this does not establish continuous updating of every household or uniform performance at local access points. The public record shows a registry, a targeting threshold and biometric authentication. It does not show a neural network deciding eligibility. Poverty measurement and thresholding can be algorithmic in a broad legal sense without being AI. Fingerprint matching, meanwhile, verifies identity; it does not determine poverty status.

We can then distinguish the possible errors. An omitted household, wrongly entered asset or outdated income indicator is a source-record problem. Linking a Computerised National Identity Card (CNIC) to the wrong household is a matching problem. A poverty formula that mismeasures urban or disability-related need presents a design or measurement problem. Fingerprint failure may arise from capture quality, occupation, device conditions, the threshold or enrolment data. Refused assistance, an unavailable office and a misleading SMS affect administrative access, while device and connectivity breakdowns may concern the vendor's payment channel. A technically accurate output

can still cause unlawful exclusion. That can happen when the underlying record is wrong or a failed authentication is treated as proof of ineligibility.

Safeguards must reach each of these stages. BISP's dynamic-survey desks, tehsil offices, helpline and 8171 channel provide identifiable routes for complaints and correction. Their existence alone does not establish accessibility, timely handling or reconsideration on the merits in every case (BISP, n.d.-e; World Bank, 2021). Notice should explain whether the problem concerns a threshold, household field, NADRA verification, duplicate or match flag, payment authentication or programme rule. Without that explanation, a woman may be sent repeatedly between BISP, NADRA and a bank. The reviewer needs authority to correct the relevant record, arrange alternative authentication where law and programme rules allow, preserve payment during urgent correction and reconsider affected downstream decisions. We must also avoid attributing every outcome to biometric matching. Its introduction coincided with changes in payment locations and agent arrangements (Clark et al., 2022). BISP remains responsible for eligibility and programme administration; the originating custodian for source records; NADRA for identity records; and regulated payment institutions for their part in disbursement.

Evidence is also needed before describing other public-service portals as automated decision systems. An online application, one-time password (OTP), searchable register or dashboard may improve access without automating a consequential judgement. Where another programme reuses a shared identifier or NSER field, however, that institution must identify its own authority, decision rule, notice and appeal. The words 'Verified by NADRA' establish no legal finding about residence, relationship, entitlement or fraud merely through identity authentication.

Policing, surveillance and criminal administration

Punjab provides the best documented policing examples, although official use of the term AI leaves much to be examined. Punjab Police reported launching a facial-recognition system in July 2023 and described further integration with CCTV and other sources in February 2024. Safe City reports also describe number-plate recognition, facial tracing and an AI-based e-challan function for specified traffic violations. Later announcements extended the infrastructure to named districts (Punjab Police, 2023, 2024a, 2024b). Reports hosted by Punjab Police describe these uses as operational; the February 2024 item reproduces reporting by The Nation. Confidence is moderate for launch and institutional use, but low for model version, camera coverage, watchlist composition, threshold, subgroup performance and independently verified effectiveness. These records establish neither nationwide deployment nor a uniform system across Pakistan.

The Punjab Safe Cities Authority Act 2016 provides an institutional basis for integrated city command, control and communications infrastructure. A model output

does not acquire police authority through that Act alone. Facial recognition may suggest a candidate match, automatic number-plate recognition (ANPR) may link a plate to a record, and traffic analytics may flag an apparent offence. Any stop, search, investigation, arrest, prosecution or use of evidence must still rest on the police and criminal-procedure powers applicable to the act. A score or ranked candidate list supports intelligence work. It is not proof of identity or guilt. The public documents reviewed did not establish an operational predictive-policing or individual criminal-risk-scoring system in Pakistan with sufficient specificity. Those functions remain unverified. Cameras and command centres do not establish them.

The possible failures again differ. Poor or obscured images concern the input. An outdated or overly broad watchlist concerns the record and its purpose. Differences in error rates across demographic groups or environments concern validation, while an incorrect candidate ranking concerns the model's output. An officer who fails to corroborate a match creates a problem of professional reliance. Proprietary systems can add another difficulty. If the institution cannot obtain model versions, comparison images, candidate lists, thresholds, audit logs or update histories, procurement has weakened its ability to produce evidence. Hong's study does not measure facial-recognition accuracy. It shows how proprietary technical, financial and knowledge arrangements can shape Pakistan's Safe City infrastructure and divide the available scrutiny (Hong, 2022).

Safeguards must match the consequence. Unseen analytics used to allocate cameras raise questions of authority, proportionality and bias. An alert initiating covert investigation requires recorded human authorisation, while a roadside intervention needs prompt verification. Material used in proceedings requires provenance, reproducibility and enough disclosure for challenge. At a minimum, records should retain the original image or event, lawful basis of the watchlist, search parameters, model and database versions, the full candidate output, operator actions, corroboration and final grounds for action. Keeping only the selected face is insufficient. People affected need routes to correct identity or vehicle records and challenge the official act. Genuine security concerns may justify controlled disclosure. They cannot make an unrecorded vendor assertion reliable evidence.

Revenue, customs, licensing, land and municipal administration

Federal revenue administration combines enacted automation powers with documented production systems. Section 175B of the Income Tax Ordinance 2001, in the official text amended up to 31 July 2025, authorises NADRA to compute indicative income and tax liability through artificial intelligence, mathematical or statistical modelling, or other methods. It provides for FBR notification to the person, an option to pay on prescribed terms, and further action under the Ordinance where payment is not made (s.

175B(4)–(8)). This is enacted authority. The legislative text does not establish that a particular model was operating for a particular population within the documentary period reviewed.

Other records provide firmer evidence of operation. FBR's Revenue Division Year Book 2024–25 reports that its machine-learning Compliance Risk Management system was operating in all Large, Medium and Corporate Tax Offices. Risk ratings and taxpayer profiles were available to assessing officers through a dashboard (FBR, 2025b, p. 37). The record supports reported operational use in those offices. It does not provide model versions, validation results or false-positive rates. Confidence is high in the official report of use, while effectiveness and the consequences of individual selections require separate evidence.

WeBOC operates nationwide as an automated customs-clearance system, assigning declarations to green, yellow and red channels and providing online staff processes, decision-support tools and activity logs (FBR, n.d.). Pakistan Customs' 2025 presentation records implementation of the machine-learning Import Risk Evaluation Engine for red-channel marking. It describes the enhanced Risk Management System (RMS) 2.0 as a simulation pilot running from 1 September 2025, with live deployment to follow testing; the printed target date is internally inconsistent (Pakistan Customs, 2025). The presentation therefore supports reported IREE use and an RMS 2.0 pilot, rather than confirmed production deployment of the latter. These processes affect importers, agents and consignments routed to scrutiny. A risk score should prompt examination. It cannot replace statutory assessment, hearing or proof.

Faceless Customs Assessment (FCA) shows the continuing human role. FBR announced implementation at Karachi from 15 December 2024 under Customs General Order No. 6 of 2024. Goods declarations from Karachi appraisement collectorates were routed to a Central Appraising Unit staffed by posted appraisers. The announcement documents this geographically bounded implementation (FBR, 2024). Its reported implementation at Karachi carries moderate status confidence, but central allocation and electronic assessment do not necessarily involve AI. FBR also reported suspending clearing-agent licences, arresting three people and opening an inquiry after an attempt to game FCA (FBR, 2025a). The episode illustrates how insider conduct, declarations and controls can undermine a system independently of model accuracy. Licence loss and enforcement still require human authority, recorded grounds and the applicable review route.

Production monitoring provides another statutory example. Section 175C of the Income Tax Ordinance 2001, inserted by the Finance Act 2025, permits posting Inland Revenue officers or other officials under the Board's control at business premises to monitor production, supply, services and unsold stock for determining tax payable. This establishes a monitoring power, not the deployment of an AI camera or automated

counting system. If software supports such monitoring, the institution must preserve the underlying observations, recording conditions, processing method and human assessment. An electronic count alone does not explain a discrepancy or determine its tax consequence.

Land administration provides a useful boundary case. Punjab's Land Records Management Information System and Arazi Record Centres operate under the Punjab Land Records Authority, offering Fard, mutation, ownership verification and record-correction services (Punjab Land Records Authority [PLRA], n.d.-a, n.d.-b). World Bank completion reporting documents digitisation and service centres across Punjab. The subsequent PULSE project supports cadastral mapping and integrated geospatial records (World Bank, 2017, 2023). Confidence is high for digital land-record operations in Punjab, while an automated title, valuation or risk model remains unverified. A wrong owner name or boundary may originate in the source or its digitisation. Failure to match a CNIC with a parcel concerns interoperability. An official who treats a screen entry as conclusive despite a pending correction raises an administrative-law problem. Electronic records do not make each of these failures model bias.

Municipal sanitation offers a bounded example of digital worker monitoring. In May 2025, the Punjab Information Technology Board (PITB) documented live facial-attendance marking through the Suthra Punjab mobile application in Faisalabad union councils and synchronisation with the waste-management company's dashboard (PITB, 2025). This supports reported use at the observed sites. It does not establish automated waste detection, province-wide model coverage or measured accuracy. Workers and supervisors need access to the relevant attendance record and a way to correct a failed or inaccurate entry. A facial-attendance result may be affected by capture conditions, identity records or the matching process. Any consequence for pay or discipline requires a competent decision under the applicable employment or contractual rules. Digital monitoring must remain open to human validation and challenge. Comparable operational AI for municipal licensing or inspection was not established in the reviewed earlier record.

Professional authority in health and education

Pakistan's health literature documents experiments, rather than a single national clinical-AI service. A mixed-methods study of 51 digital-health projects reported that 16 used AI or machine learning, 17 technologies were in a pilot phase, and most interventions were being evaluated. Only one project was linked to a national database. The study also identified gaps in skills, connectivity, cost, privacy, consent and ethics review (Kazi et al., 2020). These findings describe a mixed landscape of research, pilot and operational projects in 2020; they do not isolate the maturity of every AI/ML

project. They do not establish continuing operation beyond the study period or results applicable across Pakistan. A call in the underlying source material for interoperable electronic health records likewise expresses a need. It does not establish procurement or deployment of predictive clinical support.

The distinction matters in law. An electronic record may be incomplete without a model having failed. A diagnostic classifier may perform well on its validation data yet be unsafe in a hospital with a different language or disease prevalence. A clinician may also misuse a probabilistic recommendation. Evidence for a named clinical deployment should identify the facility, patients, intended use, training and validation populations, and professional user. It should also identify the medical-device or procurement pathway where applicable, the override procedure, adverse-event reporting and access to a second opinion. The authorised professional and provider retain clinical authority. By the cut-off, no sufficiently documented nationwide AI system for diagnosis, triage, public-health allocation or benefit prioritisation had been identified. These uses remain unverified.

Education calls for the same care. Punjab's Online College Admissions System operates as a digital application channel, and HEC's anti-plagiarism materials provide for electronic similarity reports (Higher Education Commission [HEC], n.d.; Punjab Information Technology Board [PITB], n.d.). Neither establishes automated decisions on the merits. A similarity score shows textual overlap. Academic judgement is still needed before finding plagiarism. An admissions portal may simply apply published eligibility rules without machine learning. The evidence reviewed did not establish operational AI deciding admissions, grades, student-risk classifications or educational resource allocation. These functions remain unverified. If introduced, they would require, at minimum, published criteria, validation across languages and institutions, disclosure of material records, and an authorised academic reviewer able to depart from the score. An appeal should correct the individual result and any recurring defect in the rule or dataset.

Finance, credit, insurance and fraud controls

Pakistan's credit-information system makes the allocation of responsibility relatively clear. SBP's eCIB operates nationwide for member financial institutions, with its current version and monthly reporting arrangements documented in SBP circulars. Banks and other members obtain a credit information report (CIR) and give it due weight in lending decisions. SBP expressly states that the report assigns no credit rating and creates no legal bar to lending. The financial institution decides under its own policy. Where a record is wrong, eCIB directs correction to the reporting institution, which is responsible for accuracy. Unresolved complaints can then go to SBP (State Bank of Pakistan [SBP],

n.d., 2024). Confidence is high for this record-sharing and review process. It does not show that every bank uses AI credit scoring or reveal any proprietary score a lender adds.

We need these distinctions to identify the cause of a refusal. A loan may be denied because an overdue was falsely reported, identities were mismatched, or the institution applied a particular policy. An affordability inference, staff judgement or vendor rule may also be decisive. A generic reference to a credit score conceals who holds the record and who can reconsider the result. A lender should disclose decisive factors to the extent required by law and provide a route that reaches its own underwriting process and the source institution. Where proprietary scoring materially affects consumers, the regulator needs secure access to model documents, validation and subgroup outcomes. Correcting eCIB alone will not correct an adverse inference generated by a model trained on an older copy.

SECP standards and reported disbursements establish digital lending by non-bank finance companies (NBFCs) as an operational regulated delivery model. Public material generally does not verify the operation of a particular automated credit model. SECP's Circular 15 of 2022 sets requirements for disclosure, data access, recovery conduct, complaints and credit-bureau reporting. Later materials describe regulatory and market developments in technology-mediated lending (Securities and Exchange Commission of Pakistan [SECP], 2022, 2023, 2024). Applicable requirements continue to govern the regulated entity when a software supplier provides onboarding, alternative-data analysis or scoring. Phone permissions, contacts and behavioural traces must not quietly become proxies for income, gender, location or social network without a lawful purpose and validation. Human approval offers little protection if staff cannot see the inputs, explain the score or reverse a refusal.

Banks and insurers advertise digital onboarding, fraud controls and personalised products. Such advertising does not establish a Pakistan-specific neural fraud detector, automated underwriting model, dynamic pricing engine or claims-denial system. Without records identifying the owner, version and use, these functions remain unverified. Sectoral technology-risk and consumer-protection rules may still apply if they are deployed. Vendor capabilities and foreign product documents cannot fill the gap in evidence of local operation.

Employment, platform management and consumer-facing digital markets

Platform work provides evidence of operation within a defined place and period. Fairwork's 2023 Pakistan study assessed six ride-hailing, delivery and domestic-service platforms in Islamabad and Rawalpindi. It documented app-based work allocation, management, pay, conditions, communication and appeals, and found shortcomings across several fair-work measures (Fairwork, 2023; Centre for Labour Research, n.d.). This supports operational status, with moderate confidence, for platform management

in the sampled twin-city markets at that time. It supports no nationwide claim about every worker, current model or platform. The precise ranking, pricing, deactivation and fraud-detection algorithms, together with their vendors and versions, remained largely inaccessible.

For a worker, route or task allocation affects earnings. An identity or fraud flag may lock the worker out, while an acceptance-rate measure may discipline conduct. Errors may begin in GPS or identity records, model inferences, customer ratings, hidden rules or inadequate support. A complaint button provides no meaningful review if it returns a script or reaches nobody authorised to restore access and back pay. Safeguards should include notice of the operative rule, access to material trip or account records and authorised human review. Changes to terms and scoring logic should be preserved, with regulator or court access under the applicable law.

Public evidence for named AI recruitment-screening, ranking or workplace-discipline systems in Pakistan was incomplete. Advertising an AI applicant-tracking service establishes a marketed capability. It does not establish employer deployment, the affected population or an outcome. Such uses remain unverified without supporting contracts, employer records or audits. The Faisalabad Suthra Punjab example documents facial attendance in a public-service setting (PITB, 2025). Its possible employment consequences therefore require scrutiny alongside service delivery, without inferring province-wide automated discipline from a local monitoring report.

Consumer-facing platforms operating in Pakistan do not, by their presence alone, establish a local recommendation, advertising, pricing or moderation model. The Prevention of Electronic Crimes Act 2016 (PECA), amended by the Prevention of Electronic Crimes (Amendment) Act 2025, provides for a Social Media Protection and Regulatory Authority and Tribunal. Sections 50A and 55 preserve transitional functions and earlier instruments within their terms and only so far as consistent (Pakistan Telecommunication Authority [PTA], 2021; Prevention of Electronic Crimes (Amendment) Act, 2025). The record for a challenged act must identify whether the actor was a platform, PTA within the statutory transition, or an established body under the amended law. Enactment alone does not establish a new body's operation. Pakistan-specific automated targeting, recommendation, pricing and moderation models remain unverified at system level unless linked to a named record.

Pakistan Single Window's Khizer chatbot offers a limited informational example. Its July–September 2025 newsletter reports an AI assistant on the PSW and Trade Information Portal websites for trade-related questions (Pakistan Single Window [PSW], 2025a). The October newsletter identifies Sybrid Private Limited as a development collaborator (PSW, 2025b). These institutional reports establish reported availability and a named supplier relationship, rather than independently measured accuracy or complaint outcomes. An informational reply must not be confused with a customs

determination. If a chatbot also sorts, closes or prioritises complaints, separate evidence is needed of that function, its authority and escalation rules.

Minimum safeguards across sectors

Pakistan has consequential automated processes in benefits, policing, revenue, credit and platform work. Many involve matching, rules, risk selection or digital workflows rather than neural AI. Their consequences determine the legal concern. Operation alone establishes no assurance of validation; a tender may produce no contract, and a statutory power may remain unused. An effective remedy must reach the source record, policy, contract, model or inaccessible procedure that produced the harm. The word algorithm cannot identify that cause for us.

Across these sectors, the deployer must identify its authority and responsible official, preserve the provenance of data and the versions, thresholds, outputs and interventions used, and tell people when automation and particular data materially affect a decision. It must allow source correction and actual reconsideration. Validation must address the relevant population and context, while regulators, auditors and courts need access. Errors, overrides and complaints require monitoring, with clear arrangements for suspension and correction of downstream effects. Professional supervision is meaningful only when the reviewer has the information, time, competence and authority to depart from the output. The following matrices record the available evidence and its limits.

System status in the reviewed record through 2025

The following labels state the strongest status supported by the public record for a particular system, function, institution, place and period. They establish neither accuracy nor legality, effectiveness or national coverage. 'No verified instance located' describes the limits of the documents reviewed. It does not prove that no instance exists.

Table 4.1 | System status at the manuscript cut-off

Status	Evidential threshold	Pakistan examples in sources through 2025	Source and confidence
Announced	A competent institution states an intention or launch, but the record does not establish completed procurement, a bounded trial or continuing production use.	PSW reported the launch of Khizer, an informational trade chatbot; independent performance evidence was not established.	PSW (2025a), institutional newsletter. High confidence in the reported launch; limited performance assurance.

Status	Evidential threshold	Pakistan examples in sources through 2025	Source and confidence
Proposed	A documented design, enabling provision, research plan or solicitation exists, but use is not shown. A tender notice alone is "proposed tender issued".	Section 175B permits NADRA to compute indicative income and tax liability using AI or other specified methods; this legal text does not establish actual model deployment.	Income Tax Ordinance 2001, amended up to 31 July 2025, s. 175B. High legal-authority; low operational confidence.
Procured	An award, executed contract or equivalent commitment identifies the procuring body, supplier and system.	No verified instance located in the reviewed sources that both met this threshold and identified a consequential AI function; procurement notices and project descriptions were not treated as awards.	Chapter documentary review to cut-off. Low confidence as negative evidence.
Piloted	A time-, site- or population-bounded trial is documented.	RMS 2.0 was described by Pakistan Customs as running in a simulation environment from September 2025.	Pakistan Customs (2025), official presentation. Moderate historical confidence.
Operational	Credible evidence shows use in a production process; the label does not imply complete coverage, satisfactory performance or AI in every workflow component.	BISP's NSER / 8171 / biometric-payment chain, legacy WeBOC/IREE and SBP's eCIB are documented production infrastructures. Their decision functions and territorial limits remain system-specific.	BISP (n.d.-a, n.d.-b, n.d.-c, n.d.-d, n.d.-e) and World Bank (2024), programme records; FBR (n.d.) and Pakistan Customs (2025), official records; SBP (n.d., 2024), FAQ/circular. High status confidence for the identified functions.
Suspended	A competent actor records a temporary stop in the system or function.	No verified system-wide suspension located. FBR's 2025 suspension of clearing-agent licences after an attempt to game FCA was enforcement against actors, not suspension of FCA.	FBR (2025a), official incident announcement. High confidence in the distinction; low confidence as negative system-wide evidence.
Discontinued	A competent actor records permanent termination or replacement of the identified system or function.	No verified instance located in the reviewed sources. Replacement, migration or an inactive webpage was not treated as discontinuance without an authoritative record.	Chapter documentary review to cut-off. Low confidence as negative evidence.

Status	Evidential threshold	Pakistan examples in sources through 2025	Source and confidence
Unverified	The public record does not establish the claimed status or function, or credible sources conflict. The label is not a finding of non-existence.	Named Pakistan-specific systems for several subsidy, insurance, recruitment, advertising, recommendation, pricing and automated-complaint functions were likewise not established.	Chapter documentary review to cut-off. Function-specific negative evidence remains bounded.

Table 4.2 | Pakistan-specific system evidence matrix

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
BISP NSER, 8171 and Kafaalat biometrics. Geography: federal programme / national reach. Institution: BISP, NADRA and payment partners. Population: applicant and beneficiary households, especially women recipients.	Operational. BISP (n.d.-a, n.d.-b, n.d.-c, n.d.-d, n.d.-e), official programme / service pages; World Bank (2024), programme paper; Clark et al. (2022), evaluation. High status; moderate assurance.	BISP Act 2010 and programme instruments; NADRA identity framework; BISP retains eligibility responsibility, while payment institutions remain regulated for disbursement.	Household and socio-economic records, CNIC query, eligibility threshold and fingerprint matching; identity authentication is distinct from entitlement. Degree: rule / threshold-based eligibility plus biometric identity authentication; any AI component is unverified.	Error evidence: BISP acknowledges inclusion / exclusion problems in a static registry; evaluation cannot isolate biometrics from simultaneous payment-channel changes. Outcome: 35 million household records stated; local accessibility and performance incomplete.	Human: BISP decides eligibility; NADRA corrects identity data; banks/agents disburse. Vendor: device, connectivity and payment suppliers support channels; their identities / duties are incompletely public.	Identify whether the cause is household data, threshold, match or payment authentication; dynamic survey/tehsil desks, 8171, helpline and complaint routes; human reconsideration and layer-specific correction.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
Facial recognition, Safe City analytics and e-challan. Geography: named Punjab institutions / projects, not Pakistan-wide. Institution: Punjab Police/PSCA. Population: searched persons, drivers and road users.	Reported operational. Punjab Police (2023, 2024a, 2024b), launch/project reports hosted on its website, including reproduced media reporting. Moderate status; low performance assurance.	Punjab Safe Cities Authority Act 2016 and applicable police, traffic and criminal-procedure powers; police and courts review coercive action.	CCTV images, identity / watchlist records, number plates and violation events; candidate matching, tracing and traffic flags. Degree: candidate recommendation and event-flag decision support; final coercive action remains human.	Error evidence: no public model version, threshold, subgroup error or independent effectiveness study located. Outcome: launch/use at identified Punjab sites reported; predictive policing and individual criminal-risk scoring remain unverified.	Human: authorised officers must corroborate and decide stop, search, arrest, penalty or evidential use. Vendor: proprietary suppliers / integrators are not fully documented.	Preserve source image/event, lawful watchlist basis, full candidate list, versions and officer action; give record-correction and criminal/traffic challenge routes; controlled disclosure may protect genuine security interests.
FBR machine-learning compliance-risk selection. Geography: federal tax administration. Institution: FBR. Population: cases handled by Large, Medium and Corporate Tax Offices.	Reported operational. FBR (2025b), Revenue Division Year *Book 2024–25, p. 37*. High reported-use confidence; limited performance assurance.	Tax statutes and FBR administration; competent tax authorities retain assessment and enforcement powers, subject to statutory appeal.	Returns and compliance data; risk selection for outreach, audit, scrutiny or enforcement. Degree: machine-learning risk-scoring decision support; final statutory action remains human.	Error evidence: no model versions, validation or false-positive rates supplied. Outcome: risk-rating and profiling dashboard reported in the named tax offices; causal effectiveness not independently established.	Human: FBR officials assess, hear and enforce. Vendor: supplier / model-developmen t role not publicly identified.	Tax-law notice, hearing and appeal remain necessary; disclose material discrepancies and preserve model/version, threshold, selection and override records for controlled review.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
WeBOC/IREE and RMS 2.0. Geography: federal customs/national ports. Institution: Pakistan Customs/FBR. Population: importers, exporters, agents and consignments.	Legacy WeBOC/IREE reported operational; RMS 2.0 described as a simulation pilot. FBR (n.d.), official service page; Pakistan Customs (2025), institutional presentation. High legacy-status confidence; moderate pilot-status confidence; limited performance assurance.	Customs Act and rules; FBR/Pakistan Customs; customs adjudication and statutory appeals.	Declarations, prices, history and trader profile; green/yellow/red routing, risk marking and inspection selectivity. Degree: mixed rules and machine-learning risk routing; final scrutiny and adjudication remain human.	Error evidence: model/version and false-routing data unavailable. Outcome: IREE implementation and an RMS 2.0 simulation were reported in the 2025 presentation.	Human: customs officers assess, inspect and adjudicate. Vendor: suppliers / development partners partly documented, with no complete public allocation of duties.	Preserve declaration, rule/model version, route, alert and officer action; notify material grounds through customs process; online adjudication and statutory appeal must remain available.
Faceless Customs Assessment (FCA). Geography: Karachi appraisement collectorates. Institution: FBR/Pakistan Customs and Central Appraising Unit. Population: traders, agents and consignments routed through the documented workflow.	Officially reported as implemented at Karachi, not established nationwide and not necessarily AI. FBR (2024), implementation announcement; FBR (2025a), incident report. Moderate status confidence.	Customs General Order No. 6 of 2024 and customs law; FBR/Pakistan Customs; customs adjudication and appeal.	Goods declarations and electronic allocation to posted appraisers. Degree: digital allocation and assessment workflow; any AI component is unverified, and appraisal remains human.	Error/incident evidence: FBR reported an attempt to game FCA, licence suspensions, arrests and inquiry. Outcome: implementation at Karachi reported; accuracy or comparative performance not independently shown.	Human: posted appraisers assess; competent officers decide licences / enforcement. Vendor: infrastructure or integrator role not publicly specified.	Record allocation, declarations, appraiser action, anomaly/incident records and grounds; licence or assessment consequences require notice, human authority and the applicable customs review.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
Income Tax Ordinance s. 175B mechanism. Geography: federal. Institution: FBR/NADRA. Population: persons covered by the statutory information-sharing and indicative-liability provisions.	Enabled by statute; model operation not established. *Income Tax Ordinance 2001*, amended up to 31 July 2025. High legal-authority; low operational confidence.	Section 175B; FBR/NADRA within assigned functions; applicable statutory tax review.	NADRA records and information; computation of indicative income and tax liability. Degree: AI, statistical or other computation is enabled; actual model deployment is unverified.	Error evidence / outcome: no production owner, model/version, use population, validation or individual outcome publicly documented. Enacted authority does not prove deployment.	Human: competent tax authority retains notice, assessment and enforcement. Vendor: none identified in public record.	Statutory notice/option and ordinary tax review require system-specific mapping; preserve computation, source fields and any human adoption or departure.
Statutory production monitoring. Geography: federal tax administration. Institution: FBR/Inland Revenue. Population: businesses subject to monitoring under section 175C.	Enabled by statute; AI deployment not established. Income Tax Ordinance 2001, s. 175C, inserted by Finance Act 2025. High legal-authority; low model-operation confidence.	Section 175C and applicable tax law; monitoring is directed by the competent tax authority, with tax review for consequent official action.	Production, supply, services and unsold stock; observation by posted officials. Degree: statutory human monitoring; automated sensing, counting or anomaly detection is not established by the provision.	Error evidence: no identified AI field results. Outcome: a monitoring power was enacted; this does not establish any particular sensor, model or production deployment.	Human: authorised tax officials monitor and take statutory action. Vendor: no supplier or AI contract identified by the legal provision.	Preserve source observations, recording conditions and human findings; consequential action requires recorded grounds and applicable tax review. Any supporting technology should preserve evidence for challenge.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
Punjab land records and PULSE. Geography: Punjab. Institution: PLRA/service centres. Population: landowners, transferees and service users.	PLRA digital services operational; PULSE implementation documented; AI title/valuation unverified. PLRA (n.d.-a, n.d.-b), official pages; World Bank (2017, 2023), project reports. High digital-service; low AI confidence.	Punjab Land Records Authority Act 2017 and land laws; PLRA and competent revenue/judicial forums.	Ownership, mutations, CNIC/biometric and cadastral / geospatial records; service workflow and matching, not proven AI merits decisions. Degree: digital record and matching workflow; consequential AI automation is unverified.	Error evidence: no AI performance evidence; wrong names/boundaries or mismatches are source, digitisation or interoperability errors. Outcome: digitisation and service centres documented.	Human: PLRA / service-centre and revenue officials maintain records and take legal acts. Vendor: technology providers support infrastructure; specific AI supplier unverified.	Fard, mutation and record-correction channels; reasons and adjudication remain institutional; explain the source record and legal basis rather than a generic "system" result.
Suthra Punjab facial attendance. Geography: observed union councils in Faisalabad. Institution: PITB/Faisalabad Waste Management Company. Population: sanitation workers and supervisors.	Reported operational at observed sites. PITB (2025), field-visit report dated 20 May. Moderate bounded status confidence; low model-performance assurance.	Specific enabling and consequence-specific authority was not established; any service, payment or disciplinary act must be traced to the competent body and the applicable employment or contract instrument.	Facial-attendance records synchronised with a dashboard. Degree: digital attendance and monitoring; architecture and degree of automated matching are not specified.	Error evidence: no published model validation or error rates. Outcome: field use and staff training reported; inaccurate attendance entries could affect pay or discipline if relied upon.	Human: supervisors and company officials verify attendance and decide consequences. Technology: PITB developed the application; other supplier duties are not established.	Workers need access to attendance evidence, correction and authorised challenge before adverse employment action; an attendance entry cannot replace the responsible official's reasons.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
Digital-health AI/ML projects surveyed in 2020. Geography: bounded project sites, not nationwide. Institution: participating researchers / providers. Population: project patients and professionals.	AI/ML use reported in 2020; individual project maturity and current status unverified. Kazi et al. (2020), peer-reviewed mixed-methods study. Moderate historical; low current confidence.	Health, professional, institutional ethics and procurement authorities applicable to each provider; no single national AI authority inferred.	Project-specific clinical / public-health data; diverse diagnostic, monitoring or service functions. Degree: varied decision support or monitoring; the project-specific degree is unreported.	Error evidence: study identified uneven evaluation, skills, connectivity, privacy and ethics capacity; no basis for generalisation beyond the study period. Outcome: 16 of 51 projects reported AI/ML.	Human: authorised clinicians / providers retain clinical authority. Vendor: varied or unreported across projects.	Facility-specific consent, intelligible clinical reasons, adverse-event reporting, second opinion and professional / institutional complaint route; preserve training / validation population and override.
Education portals and HEC similarity reports. Geography: identified Punjab admissions channel and higher-education institutions. Institution: PITB/colleges and HEC / institutions. Population: applicants, students and researchers.	Digital channel / similarity reporting operational; AI admissions, grading or discipline unverified. PITB (n.d.), official portal; HEC (n.d.), policy document. High channel; low AI confidence.	Published admission and academic rules; institutional / HEC oversight and available appeal / disciplinary review.	Application records and text-comparison reports; workflow / similarity support, not proven automated merits adjudication. Degree: digital workflow and similarity recommendation; merits decisions remain human.	Error evidence / outcome: no named AI merits model or performance evidence located. Similarity can be misread without context; digital operation alone does not prove AI decision-making.	Human: admission bodies and academic committees decide. Vendor: software supplies portal/report; vendor identity/version not fully documented.	Publish criteria; disclose relied-on record/report; contextual human assessment and institutional appeal; a similarity score is not itself a finding of plagiarism.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
SBP eCIB and lender decisions. Geography: national for member financial institutions. Institution: SBP and reporting / lending institutions. Population: borrowers.	eCIB operational; lender-specific AI scoring unverified. SBP (n.d.), official FAQ; SBP (2024), circular. High eCIB; low proprietary-model confidence.	Credit-bureau and banking regulation; SBP supervision; lender retains decision responsibility.	Monthly credit records and credit information report; any additional underwriting score is institution-specific and undocumented. Degree: record-sharing decision support; lender-specific model and automation are unverified.	Error evidence: SBP assigns correction responsibility to the reporting institution; no public lender-model error rates. Outcome: operational information-sharing chain documented; CIR neither rates nor legally bars lending.	Human: lender decides and reporting institution corrects; SBP supervises. Vendor: scoring/bureau suppliers beyond the documented chain are undisclosed.	Give material lending grounds where required; seek correction from reporting institution, then SBP helpdesk / complaint route; preserve the score and any human override if used.
NBFC digital lending. Geography: Pakistan-regulated channel. Institution: licensed NBFCs/SECP. Population: digital borrowers.	Delivery model operational; particular automated scoring unverified. SECP (2022), regulatory circular; SECP (2023, 2024), official releases. High channel; low model confidence.	SECP digital-lending requirements and NBFC law; SECP regulates the licensed entity.	Application, bureau and permitted device/account data; onboarding, credit assessment, pricing and recovery. Degree: technology-mediated lending with limited human intervention is regulated; the particular scoring degree is unknown.	Error evidence: no public model/version or group-error results. Outcome: rules and disbursement activity documented, not an individual model's operation or effects.	Human: licensed NBFC retains the lending decision and complaint duty. Vendor: platform/scoring supplier may support, but contracts/models are not public.	Key disclosures, data limits, recorded reasons where applicable, complaint and bureau-correction routes; human reviewer must see inputs and be able to reverse a decline.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
Platform work. Geography: Islamabad / Rawalpindi sample, 2023. Institution: six sampled platforms. Population: drivers, couriers and domestic-service workers.	App-mediated management operational in the bounded sample. Fairwork (2023), research report; Centre for Labour Research (n.d.), programme page. Moderate bounded confidence; no nationwide inference.	Contract, labour and other applicable law; forum depends on legal character of relationship and claim.	Location, availability, ratings, acceptance, task and payment data; allocation, pay, discipline, deactivation and fraud flags. Degree: app-mediated ranking, allocation and management; exact automation and human override are unknown.	Error/outcome evidence: sampled platforms fell short on several fair-work measures; exact models, versions and error rates inaccessible. Evidence does not establish every platform or current practice.	Human: platform personnel set rules and should reconsider lockout/pay decisions. Vendor: algorithm/vendor chain opaque.	Notice of operative rule; access to trip/account evidence; authorised human appeal able to restore access/back pay; preserve terms and scoring changes for regulator/court review.
Social-media moderation and public complaint / removal route. Geography: Pakistan-facing platforms. Institution: platforms and competent statutory bodies, including PTA where the transition applies. Population: users, speakers, audiences and complainants.	Pakistan-specific moderation model/version is unverified. PTA (2021), rules; PECA 2016 as amended in 2025, including the transitional provisions. High confidence in the statutory text; low model confidence.	The amended Act provides for establishment of the Social Media Protection and Regulatory Authority and Tribunal; ss. 50A and 55 govern transition and continuity. The actor's authority must be verified for the date of the challenged act.	User content, reports and platform signals; detection, ranking, removal, blocking and account restriction may be human, automated or mixed. Degree: mixed or unknown moderation; final statutory action remains attributable to an authorised human body.	Error evidence / outcome: individual automation role, input, version and rationale usually opaque; service presence and legal route do not prove model accuracy or automated selection.	Human: platforms and competent statutory bodies act within their powers; courts retain review. Vendor: platform model/vendor chain generally undisclosed.	Platform appeal and the applicable statutory complaint/review route; constitutional or other judicial review as available. Identify the competent body, provide item-specific grounds and enable controlled disclosure.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
PSW Khizer informational assistant. Geography: PSW and Trade Information Portal websites. Institution: Pakistan Single Window. Population: traders and other trade-information users.	Reported launched and integrated in 2025. PSW (2025a, 2025b), institutional newsletters. High reported-launch confidence; limited independent performance evidence.	Trade-information support; an informational assistant has no independent authority to make a customs determination.	User questions and trade information; guidance on procedures, compliance and documentation. Degree: automated informational assistance; no adjudicative function is established.	Error evidence / outcome: no independently measured accuracy or complaint-outcome evidence in the cited reports. Informational availability does not establish legally reliable answers.	Human: PSW and competent trade authorities retain their official functions. Vendor: Sybrid Private Limited identified as development collaborator.	Display limitation and responsible authority; escalate users to formal complaint channels; reassess status and safeguards before any automated triage, prioritisation or closure.
Automated subsidy allocation or benefit prioritisation beyond documented BISP targeting. Geography / institution: system-specific; none established. Population: applicants and recipients of subsidies or public benefits.	Unverified. BISP programme records and World Bank (2024) document targeting infrastructure, but no owner-, version- and use-specific AI subsidy allocator was located. Low confidence as negative evidence.	Enabling statute / programme rules of the relevant benefit authority; sector ministry / department and available ombuds/judicial review.	Likely eligibility, household, income, identity and programme data; ranking, allocation or fraud detection alleged but not established. Degree: consequential automation is unverified.	Error evidence / outcome: none system-specific; do not infer a model from a registry, threshold or digital payment.	Human: benefit authority would remain responsible for entitlement / allocation. Vendor: no identified supplier or contract.	Any deployment requires adverse notice, decisive data and factors, urgent correction, authorised reconsideration, benefit restoration and downstream correction.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
Automated generation and automated complaint triage, prioritisation or closure. Geography / institution: public bodies and regulated firms; no named consequential system established. Population: complainants, service users and officials relying on generated text.	Unverified for consequential complaint decisions. PSW (2025a, 2025b) documents Khizer as informational support; this does not establish automated complaint closure. Low confidence as negative evidence.	The decision-maker's enabling statute, complaint rules and record duties; relevant regulator / ombuds / court.	Complaint text, account/service records and generated replies; summarisation, drafting, routing, prioritisation or closure. Degree: consequential generation or complaint automation is unverified.	Error evidence / outcome: no public owner/version, false-routing rate, generated-decision record or measured outcome located.	Human: authorised complaint officer must verify content and decide priority / closure. Vendor: none identified; any provider must preserve prompts, outputs, versions and logs.	Disclose automation's material role; preserve complaint, generated output and edits; accessible human escalation and reopening; generic generated text cannot replace intelligible reasons.
Insurance underwriting, pricing, fraud flags or claims denial. Geography: Pakistan-regulated insurance market. Institution: insurer/SECP. Population: applicants, policyholders and claimants.	Unverified at named-system level. Reviewed SECP digital-finance materials through the cut-off do not identify an insurer, model/version, affected population and operational outcome. Low confidence as negative evidence.	Insurance Ordinance and SECP regulation; insurer remains responsible; Insurance Ombudsman, SECP and courts as applicable.	Application, policy, health, claims, fraud and possibly behavioural data; underwriting, pricing, investigation or claims recommendation. Degree: consequential automation is unverified.	Error evidence / outcome: none system-specific; digital onboarding or personalised-product advertising does not prove an automated underwriting or denial model.	Human: authorised insurer personnel decide and investigate. Vendor: no named scoring / analytics provider established.	Give policy/legal grounds and material data; correct records, obtain merits reconsideration and use insurer, ombuds, SECP or court route; preserve score and fraud-flag evidence.

Pakistan-specific system evidence matrix: system, geography, institution and population	Status; source, date and document type; confidence	Legal authority and regulator	Data, function and degree of automation	Error evidence and documented outcome	Human role; vendor role	Notice, explanation and review route
Conventional recruitment screening, ranking or workplace discipline. Geography / institution: Pakistan employers generally; no named deployment established. Population: applicants and employees.	Unverified. Fairwork (2023) documents bounded platform management, not conventional recruitment; vendor marketing without employer contract/use evidence was excluded. Low confidence as negative evidence.	Employment, service, equality, contract and professional rules applicable to the employer; labour/service forums or courts.	CVs, tests, identity, attendance, productivity and communications; screening, ranking, monitoring or discipline. Degree: consequential automation is unverified.	Error evidence / outcome: no named employer/model, validation, group-error evidence or outcome located. Suthra Punjab is separately recorded because official sources identify monitoring.	Human: employer / appointing or disciplinary authority must decide. Vendor: no verified applicant-tracking / model provider in use.	Tell the person the decisive criteria/data and automation's role; accessible correction and authorised reconsideration; preserve vacancy criteria, score, version and override.
Targeted advertising, recommendation and dynamic pricing. Geography: Pakistan-facing digital markets; no named local model established. Institution: platforms / traders; relevant sector and competition / consumer authorities. Population: users and consumers.	Unverified at Pakistan system/version level. Fairwork (2023) and PTA (2021) concern adjacent platform practices and content procedure, not proof of a named local ad, recommendation or pricing engine. Low confidence as negative evidence.	Competition, consumer, electronic-transaction and sector law; Competition Commission, consumer forums and sector regulators as applicable.	Profiles, location, browsing, purchase and interaction data; audience targeting, content/product ranking and personalised or surge prices. Degree: consequential automation is unverified.	Error evidence / outcome: no owner/version, local validation, discriminatory pricing or documented outcome established; platform presence and foreign transparency material are insufficient.	Human: trader/platform remains responsible for offer, price and complaint handling. Vendor: ad-tech / recommender vendors and contracts unverified.	Clear commercial disclosure where legally required; explanation of material price / eligibility factors, data correction and human complaint route; preserve targeting, ranking and price-decision logs for regulator/court access.

4.2 Dignity, Informational Privacy, Equality, and Differentiated Vulnerability

Automated classification operates through institutions. A record is collected, a person or household is matched, an inference is produced, and an authorised actor decides whether to use it. The rights inquiry must follow the same sequence. Building on the data and identity arrangements in Sections 2.1–2.2, we ask whether collection, inference, disclosure, classification or action affects an interest protected by the Constitution or an applicable statute, regulation, licence or common-law duty. The absence of the term artificial intelligence from a law does not end the inquiry. Lawful collection does not authorise every inference, accurate authentication does not decide entitlement, and a useful score does not itself justify adverse treatment. Sections 4.3 and 4.4 connect these interests to disclosure, proof, forums and remedies.

Constitutional protection and the limits of analogy

The Constitution provides connected protections with distinct scopes. Article 4 gives every citizen and every other person for the time being within Pakistan an inalienable right to legal protection and treatment in accordance with law. It prohibits action detrimental to life, liberty, body, reputation or property except in accordance with law. Article 9 protects every person against deprivation of life or liberty save in accordance with law. Article 14 makes human dignity inviolable and protects privacy of the home subject to law. Article 10A guarantees fair trial and due process in determining civil rights and obligations or a criminal charge. Under Article 19A, every citizen has a right of access to information in matters of public importance, subject to regulation and reasonable restrictions imposed by law. Article 25 guarantees citizens equality before and equal protection of law, prohibits discrimination on the basis of sex and allows special provision to protect women and children (Constitution of the Islamic Republic of Pakistan, 1973, arts. 4, 9, 10A, 14, 19A, 25).

We should give these provisions their proper force without treating them as one general constitutional right to data protection. Article 14 distinguishes inviolable dignity from privacy of the home 'subject to law'. The Supreme Court has nevertheless read privacy alongside dignity, liberty, life and expression, extending the inquiry beyond a building. In Mohtarma Benazir Bhutto v President of Pakistan (PLD 1998 SC 388), telephone tapping raised a constitutional issue. It could not be dismissed as a technical telecommunications feature. Shehla Zia v WAPDA (PLD 1994 SC 693) likewise understood life under Article 9 more broadly than physical survival. These authorities support protection against serious informational intrusion according to context. They do not make privacy absolute.

The treatment of bodily and identity information helps explain the issue. In Laila Qayyum v Fawad Qayum (PLD 2019 SC 449), the Supreme Court held that forcibly taking a woman's DNA to determine paternity would violate liberty, dignity and privacy protected by Article 14. It also recognised the attempt to deprive her of identity and inheritance. This was not an automated-profiling case. It nevertheless shows that information can be inseparable from legal identity. Technical value as proof does not remove the need for authority, necessity and respect for protected liberty. This book uses 'informational autonomy' as shorthand for control over compulsory disclosure, bodily information, identity and consequential use. The expression is analytical, rather than an express constitutional term.

Article 4 expressly protects reputation, and the concern can arise without public publication. A fraud flag, security label, predicted affiliation or credibility score may damage a person's reputation within government or a regulated firm. We need to identify who produced and communicated it, whether it was treated as fact or probability, and what action followed.

Article 25 also requires a precise reading. It guarantees citizens equality before and equal protection of law, without requiring identical outcomes. The reasonable-classification doctrine permits an intelligible distinction rationally related to the lawful object, while rejecting arbitrariness (I. A. Sherwani v Government of Pakistan, 1991 SCMR 1041). Articles 26 and 27 address public places and public service on enumerated grounds. Religious profiling may engage Article 20. In Dr Muhammad Aslam Khaki v S.S.P. (Operations), Rawalpindi (PLD 2013 SC 188), the Supreme Court directed public authorities to protect khawaja sira persons' equal rights and participation, including identity and service arrangements. Classification and implementation can therefore be reviewed. A fairness metric chosen by a vendor cannot settle the constitutional question.

Statutory protection is uneven. PECA, as amended, criminalises specified unauthorised access, copying, interference, identity misuse and conduct affecting privacy or dignity. Section 16 provides a limited route for unauthorised use of identity information. PECA deals with offences and investigation; it is not a general code governing every lawful collection, profile, retention decision or automated outcome. Right-to-information laws permit access to covered public records subject to exemptions, but do not provide complete rights of data-subject access, correction, restriction and erasure (Right of Access to Information Act, 2017, ss. 4–8, 16–20). Sector laws may impose further duties. Constitutional guarantees principally govern law and public action. Private claims ordinarily need a statutory, regulatory, contractual or common-law basis. Outsourcing, however, does not turn the State's duties into private ones.

At the manuscript cut-off, the Ministry of Information Technology and Telecommunication's Personal Data Protection Bill 2023 remained a draft. The Ministry's legislation register and the National Assembly's Acts register listed no enacted personal-data-protection Act (Ministry of Information Technology and Telecommunication [MoITT], n.d.-a, 2023; National Assembly of Pakistan, n.d.). The Bill defines profiling and sensitive data and proposes rights relating to notice, correction, harmful processing, erasure, grievances and solely automated decisions. These provisions help identify gaps and possible reforms. Their appearance in an official draft does not make them enforceable. Cybersecurity and AI policies likewise cannot create coercive power or displace constitutional and statutory duties.

Machine-generated inferences as consequential information

Records governance asks whether information was lawfully collected, accurate, secure, properly disclosed, retained and corrected. Automation adds something further: a claim about the person. A transaction may be labelled anomalous, an applicant high-risk, a face similar to a watchlist image, or a household likely to be ineligible. Such an attribute may never have appeared in the source database and may be only probabilistic. It becomes legally consequential when it changes scrutiny, delay, access, investigation, price, reputation or burden.

Authority to collect information must be distinguished from authority to infer from it. Registering identity data does not itself authorise behavioural prediction, political profiling or unrelated enforcement. Authentication also has limits. A match may link a credential with its presenter, but deciding citizenship, culpability, benefit eligibility or professional fitness requires the relevant substantive rule. Finally, correlation cannot supply a legal reason on its own. The institution must assess whether a predictive factor is relevant, reliable, permissible and sufficiently probative for the task.

These distinctions become particularly important when linked records construct a person's administrative identity. Hashmi (2021) describes how genealogical computation in Pakistan turns documentary descent and family relationships into evidence of belonging and reliability. A wrong link may therefore do much more than misspell a name. Reused as a signal of status or risk, it can change how institutions recognise the person. Protection requires scrutiny of an inference's provenance, meaning, confidence, age and permitted use. It also requires the model and threshold that produced it, the recipients to whom it travelled, and a procedure for challenging source facts and evaluative judgement alike.

A rights test for surveillance, biometrics, linkage, and secondary use

Surveillance and predictive administration differ in their intrusiveness. Authentication requested for a service is different from covert interception, persistent tracking, watchlist matching and predictive policing. Each should still be examined through authority, purpose, necessity, proportionality, retention, security and unequal effects. Justification should become stronger as secrecy, scale, sensitivity, irreversibility and consequences increase.

To establish authority, we must identify the competent actor, enabling instrument, permitted data and authorised consequence. Section 54 of the Pakistan Telecommunication (Re-organization) Act 1996 authorises federal interception or tracing in its specified national-security or offence-related circumstances. The Investigation for Fair Trial Act 2013 provides a warrant-based regime for scheduled offences. PECA governs preservation, traffic-data retention, search, seizure and real-time collection within its investigative field (Investigation for Fair Trial Act, 2013; Pakistan Telecommunication (Re-organization) Act, 1996, s. 54; Prevention of Electronic Crimes Act, 2016, ss. 31–36, 39). These routes have different thresholds, scopes and procedures. A lawful-intercept interface or security label does not show that a particular operation, reuse or automated flag was authorised.

Purpose must be stated precisely enough to explain the interference with rights. Security, fraud prevention and better services do not tell us who may be monitored, what may be inferred or which action may follow. Original collection is a separate use from matching, training, watchlist comparison, investigation and use in evidence. Predicting criminal association from payment-authentication data requires its own legal basis and safeguards. Consent cannot supply statutory competence where none exists. Nor may it be voluntary when an indispensable service depends on giving it.

Necessity asks whether a measure materially advances a lawful purpose and whether a reasonably effective, less intrusive alternative exists. The ability to do something does not establish the need. Bulk collection, continuous monitoring and population-wide biometric matching require more than administrative convenience. Proportionality weighs the expected contribution against the depth, duration and scale of intrusion, errors, chilling effects and safeguards. These considerations do not form a single mechanically settled test for every Pakistani right. They organise the context-specific inquiry into authority, rational connection, reasonableness, dignity, liberty and fair process.

Retention requires separate treatment of raw records, biometric templates, authentication events, watchlist candidates, inferences, investigative files and decision records. PECA section 32 sets a minimum period for traffic data within its field. Digital Identity regulations 25 and 30 and National Data Exchange Layer (NDEL) regulation 10 require indefinite retention of specified minimum transaction data, while NDEL entity

logs ordinarily have a five-year period. These rules govern particular channels. They do not authorise keeping every profile or payload indefinitely. A lawful schedule should preserve scores and provenance needed for appeals or incidents, restrict contested or expired uses, and provide secure disposal where no continuing purpose remains.

Security protects information; it does not establish the legality of its use. Access controls, encryption, segregation, audit logs, integrity checks, supplier controls, incident response and breach handling are particularly important for biometrics, which cannot be reset like passwords. Encryption does not make an irrelevant inference relevant. Sensitive details may properly be withheld from publication while remaining accessible under controlled conditions to an authorised court, regulator or auditor. Secrecy cannot remove the need for an accountable legal basis.

NADRA's current instruments show how these duties apply to specific channels. The National Data Repository Regulations 2024 require a purpose and justification, necessity assessment, sharing terms, security, retention and disposal. Regulation 16 nevertheless dispenses with explicit consent for sharing through its section 7(i) route. The National Data Exchange Layer Regulations 2025 govern federated exchange through approved services and field limits, consent subject to stated exceptions, minimisation, security, incident response and logs. Originating and on-boarded entities remain responsible for their own data and operations. The Digital Identity Regulations 2025 require notices specific to purpose and data, constrain reuse, provide auditable authentication histories, and support security, accessibility, language choice and backup authentication (NADRA Digital Identity Regulations, 2025; NADRA National Data Exchange Layer Regulations, 2025; NADRA National Data Repository Regulations, 2024). These instruments create no universal privacy code and transfer no service agency's substantive decision power to NADRA.

Direct, indirect, proxy, and cumulative discrimination

We need to describe automated differentiation accurately. Direct discrimination occurs when an actor treats someone adversely because of a prohibited or otherwise impermissible characteristic. Using an explicit sex field to exclude women from a programme is a clear example. But using a characteristic is not invariably unlawful. Article 25(3), a targeted remedial programme, a clinically relevant fact or another valid legal distinction may justify particular treatment. The inquiry concerns authority, purpose, relevance and effect. The presence of a field in a dataset cannot settle it.

Indirect discrimination describes an apparently neutral criterion or procedure that places a distinctive burden on a group. Smartphone-only authentication, stable formal employment, a particular script or repeated visits to a registration centre may look uniform. In practice, poverty, disability, language, restrictions on women's mobility, rural residence and documentation can make those requirements fall unevenly. The

expression helps explain the problem, but it is not necessarily an independent cause of action under every Pakistani statute. Claimants and reviewers must connect the effect to Article 25, another applicable provision or a statutory duty. They must then examine the lawful object, the closeness of the criterion's connection to it and the availability of a less exclusionary alternative.

Proxy discrimination arises when an apparently neutral feature represents, correlates with or enables reconstruction of a protected or structurally disadvantaged characteristic, causing group-linked disadvantage. Postcode, school, language, device type, typing rhythm, address stability, family association, transaction location and access time may convey information about region, wealth, disability, gendered access, minority identity or migration. Deleting a protected field does not delete information distributed across other fields. At the same time, carefully governed demographic data may be needed to measure unequal error rates. Evaluation should therefore have a limited purpose, controlled access and reliable analysis of small groups. Refusing to collect any relevant information can leave disparity impossible to measure.

Cumulative discrimination concerns harm caused by repetition and linkage. An error in a source record may cause an identity mismatch. That mismatch may become an ineligibility flag, which a payment channel or another programme then reuses. Repeated failures may themselves become a signal of behavioural risk. No single institution appears to have adopted a discriminatory rule, yet the person faces lasting exclusion. Cumulative discrimination is thus a category for evidence and remedy. It directs attention to the complete sequence, shared identifier, recipients, feedback loops and actors able to stop further transmission.

Technical neutrality provides an inadequate defence against these problems. Data mining can reproduce disadvantage through the chosen target, training examples, features, proxies and thresholds without an express instruction to discriminate (Barocas & Selbst, 2016). Numerical disparity alone does not decide legality either. Analysis must identify the affected interest, relevant comparator, legal prohibition, institutional purpose, cause and appropriate remedy. Separating the model from its administrative setting can conceal the practices causing the harm (Selbst et al., 2019).

Representation, measurement, comparators, and unequal error

Representation concerns who and what appears in development and validation data. Even a large dataset may contain little evidence about rural districts, regional languages, women sharing devices, persons with disabilities, religious or ethnic minorities, migrants and people without stable documents. Intersections of these groups may be thinner still. Migration claims also need provision-specific analysis: Article 25 is citizen-specific, while Articles 4, 9 and 14 protect persons in their relevant terms. Missing data often reflect exclusion. People denied a service produce fewer successful records and may disappear

from the denominator used to claim accuracy. Later data may then seem to confirm the system's view because rejected people cannot produce the outcomes needed to test it.

Measurement asks whether a recorded variable represents the legal or professional concept that matters. Fingerprint quality does not establish identity, formal banking does not establish honesty, and continuous phone access does not establish eligibility. An address cannot settle permanent belonging, a past arrest cannot establish culpability, and historical receipt of benefits is not a flawless measure of need. Labels taken from earlier decisions may preserve discretion, unequal enforcement, inaccessible procedures and outdated records. A model can learn these patterns accurately while measuring the wrong thing. Jacobs and Wallach (2021) therefore place part of fairness in construct validity: choosing what to measure and how to represent it comes before calculating parity.

Choosing a comparator is a legal as well as a technical task. Overall accuracy may conceal a subgroup's high false-non-match rate. Equal false-positive rates may impose very different burdens where exposure and base rates differ, and group averages can conceal failures at the intersection of characteristics. As Section 3.4 explains, validation, explanation and audit must examine the deployed population and context. An aggregate benchmark is insufficient. Studies of commercial gender classification and face recognition show that demographic effects vary with algorithm, task, dataset and subgroup (Buolamwini & Gebru, 2018; Grother et al., 2019). They do not establish how a Pakistani system performs. They explain why a vendor's global average cannot prove equal reliability in Pakistan.

Local research also supports testing in context. Urdu evaluations vary substantially by task and model, making validation against the actual language, script, institution and population more useful than reliance on English benchmarks (Arif et al., 2024; Tahir et al., 2025). Social exclusion differs across regions, professions and ethnolinguistic groups even after controlling for income (Pervaiz et al., 2021). Research on Lahore also shows how securitised urban space and policing intersect with class, gender and ethnicity in shaping women's mobility (Abbas, 2024). These studies establish no algorithmic discrimination by themselves. They identify dimensions along which sampling, access, surveillance burdens and consequences should be investigated.

An evaluation of BISP's fingerprint-verification rollout found improved control of cash for women and no average decline in successful collection in the setting studied. It also recorded difficulties for people whose fingerprints were hard to read and failures involving devices or connectivity (Clark et al., 2022). We need to preserve both findings. Governance should publish denominators and failure codes, disaggregate results on a defensible basis, provide backup authentication and assisted correction, and test whether those remedies work.

Validation should report false positives and false negatives for legally and operationally relevant groups, with confidence intervals and the limits of small subgroup cells. It should account for missing and rejected cases, geographic and channel coverage, language and accessibility, threshold choice, overrides, complaints and change over time. Each metric should identify the equality proposition it informs. Statistical parity cannot supply lawful authority. Unequal rates do not automatically complete a constitutional claim. They form part of the evidence, alongside the affected interest, comparator, justification, institutional conduct and available accommodation or correction.

Cumulative exclusion and effective correction

The National Socio-Economic Registry illustrates how exclusion can spread without proving any AI deployment. BISP describes it as a household repository used for targeting and cross-validation with NADRA. An independent assessment found that dynamic registration improved inclusion, while desk-based procedures could underserve remote households. Outreach was particularly needed in Balochistan (BISP, n.d.-c; Majoka et al., 2024). A stale record, disputed family link, failed biometric, targeting rule and inference each arises at a different point and needs different evidence. A person seeking urgent assistance should not first have to identify the responsible database or contractor. The institution should route the correction.

An effective remedy must first contain immediate harm. Where delay threatens subsistence, liberty, health, education or safety, this may require alternative authentication, provisional service, restrictions on adverse reuse or an expedited human decision. The authoritative source must then be corrected while preserving the audit trail. That correction must reach recipients, and stale copies must be blocked. Affected decisions need reconsideration, including arrears, restored status and removal of flags. Finally, recurring failures require examination of the match rule, labels, threshold, model, agreement, contract, staffing or complaint process. Audit, retraining, suspension or decommissioning may be proportionate responses.

Table 4.3 | Administrative exclusion and corrective responses

Exclusion point	Protected interest and characteristic risk	Actor with primary corrective power	Urgent individual response	Systemic response
Source record or family link	Identity, reputation, status; documentation, migration, minority or gendered family pattern	Originating registrar or custodian; NADRA for its identity record	Show provenance, mark dispute, correct or give reasoned refusal	Quality audit; feeder-update control; recipient inventory

Exclusion point	Protected interest and characteristic risk	Actor with primary corrective power	Urgent individual response	Systemic response
Authentication or biometric comparison	Access, subsistence, liberty; age, disability, manual work, device or connectivity	NADRA/requesting entity and service operator	Failure code, backup modality, assisted verification	Subgroup failure testing; equipment and accessibility review
Match, profile, or risk inference	Equality, privacy, reputation; language, region, poverty, association proxies	Model deployer and substantive decision authority	Disclose material basis, restrict adverse reliance, human reconsideration	Revalidate features, labels and threshold; retrain, rollback or suspend
Downstream denial or investigation	Benefit, movement, education, finance, liberty or fair process	Institution legally empowered to decide the service or case	Restore access where justified; reconsider; provide appeal and interim relief	Trace referrals and overrides; revise rule, workflow and contract
Repeated cross-system propagation	Multiple interests; cumulative and intersectional disadvantage	Coordinating owners plus every recipient; no transfer of each body's duty	One case identifier; routed complaint; correction and retrospective relief	Joint incident review, recipient-wide correction, independent supervision

The matrix does not assign every error jointly to all actors. It makes their respective responsibilities visible. The source custodian answers for the authoritative record, and the exchange operator for secure, accurate transmission within its mandate. The model owner answers for inference and validation, while the sector authority answers for the legal decision, notice, review and remedy. Vendors may have contractual or private-law responsibility. A public body, however, cannot answer an equality or dignity challenge simply by saying that software supplied its score.

Vulnerability is not a permanent characteristic of a group. It arises through the interaction of a person's circumstances with the system's data, channel, design, location, timing and consequences. Pakistan's Constitution, identity regulations, sector laws, criminal-procedure safeguards, information-access regimes and judicial precedents already govern much of that interaction. Gaps remain, particularly the absence at the cut-off of enacted general personal-data rights covering consequential inferences and automated decisions. Those gaps support precise legislation and sector rules. They do not suspend existing duties. Before proposing a new AI-specific right, a deploying institution must be able to explain its authority, purpose, evidence, unequal effects, decision-maker, correction route and remedy. That explanation must address the individual error and the process that allowed it to spread.

4.3 Procedural Fairness, Disclosure, Machine-Generated Evidence, and AI in Legal Practice

A decision influenced by an algorithm raises two procedural questions. First, did a legally competent person apply lawful criteria, accurate material and independent judgement, while giving the affected person a fair chance to understand and challenge the result? Second, when an output, log, biometric match, generated document or allegation of synthetic media enters proceedings, can the tribunal establish its source, integrity, limits and relevance to a fact in issue? Saying that AI was used answers neither question. Technical sophistication does not relax ordinary legal standards. It makes it more important to identify the proposition being advanced, the person or institution advancing it, and the records against which it can be tested.

Pakistan's procedural baseline is dispersed but substantial. Article 4 protects treatment in accordance with law. Article 10A guarantees fair trial and due process when civil rights and obligations or criminal charges are determined. Article 14 protects dignity and, subject to law, privacy of the home. Article 19A gives citizens access to information subject to lawful regulation and reasonable restrictions, while Article 25 guarantees citizens equality before and equal protection of law (Constitution of the Islamic Republic of Pakistan, 1973, arts. 4, 10A, 14, 19A, 25). Section 24A of the General Clauses Act 1897 requires powers conferred by or under an enactment to be exercised reasonably, fairly, justly and for its purpose, with reasons so far as necessary or appropriate. The Supreme Court has connected adverse action to notice and a hearing, subject to context-specific exceptions. It has also required speaking orders to show application of mind, engagement with relevant law and material, and a rational connection between facts and conclusion (Warid Telecom (Pvt) Ltd v Pakistan Telecommunication Authority, 2015 SCMR 338; Mustafa Impex, Karachi v Government of Pakistan, PLD 2016 SC 808). Automation enters this existing procedure. It does not replace it.

Notice and intelligible reasons

Notice must allow a challenge before an adverse result becomes irreversible. Its content depends on the enabling law, forum, urgency and consequences. For a high-consequence decision assisted by AI, 'failed verification' or 'high risk' will ordinarily be inadequate. A person should receive the outcome and effective date, legal authority, operative criterion, material facts and records attributed to them, and the function performed by automation. Notice should also identify the responsible institution and authorised official, together with correction, reconsideration and appeal routes, deadlines and available interim protection. If a score, ranking, match, summary or recommendation materially influenced the decision, that role should be stated. A database mismatch

should be distinguished from a fraud prediction, and an alert from the official finding that leads to denial, investigation or sanction.

Intelligible reasons connect authority, the applicable criterion, material evidence and judgement. As Section 3.4 explains, a fraud-risk score and feature attribution explain an output. They do not explain the legal decision. Reasons must show whether predicted fraud is an authorised criterion, whether the records concern the right person, why the threshold carries the stated consequence, how contrary evidence was considered and who decided. Generating a paragraph that repeats the score does not demonstrate independent application of mind.

Reasons need not reveal source code or every parameter. The person usually needs material inputs, significant factors, an understandable account of the threshold or rule, the degree of automation and the human action taken. A material limitation, such as weak validation for a language, region or subgroup, must not disappear behind a categorical label. Disclosure still depends on context. Revealing a live fraud signature or investigative method may enable evasion, and other people's records need protection. But a security claim cannot make an unreviewable assertion a reason. At minimum, the competent decision-maker and authorised reviewer need the unredacted basis. The affected person should receive the fullest useful account consistent with a lawful restriction justified for the particular material.

The federal Right of Access to Information Act 2017 supports access without creating a complete code of algorithmic rights. It covers federal public bodies, requires proper record maintenance and indexing, and mandates proactive publication of applicable rules, criteria, standards and decision processes. Contracts, final orders and decisions are public records subject to exclusions and exemptions (ss. 1, 4–7). A refusal must identify the exact exception and appeal route, and the public body bears the burden before the Pakistan Commission on Access to Information (ss. 13, 17). The Act protects privacy, commercial confidence, system security, privilege and court proceedings, while requiring severance when only part of a record is exempt (s. 16). An RTI request is one route to information. It does not replace party disclosure, statutory appeal, constitutional review or reconsideration on the merits. Provincial regimes and sector procedures require separate examination.

Genuine human reconsideration and accessible challenge

A person at the end of a workflow does not, by their presence alone, provide meaningful review. Reconsideration needs a named reviewer with legal power to change the result, access to the complete case file and source records, and enough information about the operative model, dataset, configuration and threshold to understand the output. The reviewer also needs time, domain competence and practical power to stay, reverse, vary, restore, correct or escalate the decision. Checking that the software ran is insufficient.

The reviewer must be able to disregard its score and record their own assessment of the person's representations. Where the governing scheme provides merits review, the reviewer should reconsider the merits. An official cannot reduce that entitlement to a check of system integrity.

Confirmation becomes mechanical when officials see only a red or green flag, face incompatible production targets, cannot override the result, lack contested records or need vendor permission to understand the output. An approve button may conceal the absence of judgement. We therefore need operational evidence: review time, available information, recorded reasons, requests for further material, overrides, reversals and escalation outcomes. No overrides does not by itself prove automation bias. It is nevertheless a warning where cases differ and the reviewer is expected to exercise discretion.

Accessibility is part of a fair procedure. Notice and challenge should be available in relevant languages and accessible formats, with assisted, in-person and offline routes where distance, remoteness, connectivity, disability, literacy, cost or missing documents make digital-only access ineffective. A stable case reference must follow the person through referrals between the data custodian, system operator and deciding institution. Loss of subsistence payments, access credentials, liberty-related protection or time-sensitive treatment may require a stay, provisional service or accelerated review. A successful challenge can otherwise arrive too late. Section 10(2) of the federal RTI Act requires assistance where disability makes it difficult to describe the requested information. This illustrates one accessibility duty, without exhausting the subject. The Legal Aid and Justice Authority Act 2020 provides legal aid for poor and vulnerable persons within its defined criminal-justice mandate. It does not guarantee representation in every administrative dispute.

Preservation, disclosure and controlled technical access

Evidence needed for review must often be created before any dispute arises. A deploying institution should preserve the input snapshot and provenance, dataset, model and configuration versions, feature and label definitions, validation population and results, thresholds and uncertainty measures. The record should also retain outputs and explanations, access and event logs, human interventions, final reasons, complaints, incidents and changes. The Electronic Transactions Ordinance 2002 recognises electronic form and sets requirements concerning writing, originality and retention. These include later accessibility, accurate representation of content and form, and origin, destination, date and time where applicable (ss. 3–6). It sets no universal retention period. The National Archives Act 1993 separately governs covered federal public records, including machine-readable material. Retention schedules must protect

foreseeable complaints, appeals, audits and litigation without keeping material indefinitely for no purpose.

Records must remain available across organisational boundaries. Federal procurement within its scope is governed by the Public Procurement Regulatory Authority Ordinance 2002 and Public Procurement Rules 2004. Solicitation and contract terms should secure the evidence described here consistently with those instruments. The deploying body, regulator, auditor, court or tribunal, and appropriately bound independent experts should have access needed for their functions. Vendor ownership and intellectual property may shape confidentiality arrangements and commercial liability. They cannot veto judicial process or remove a public body's duty to justify its act. A high-consequence service should not be procured if the institution cannot preserve the operative version and reconstruct an output after an update. Wexler (2018) describes the structural conflict where proprietary criminal-justice tools prevent defendants from testing evidence affecting liberty. Pakistan must resolve that conflict through its own procedural, evidentiary, constitutional and contractual law. Procurement should secure access before deployment.

Disclosure does not always require publication. The public may need policies, aggregate performance and institutional information. The affected party needs material facts, the role of automation and a fair opportunity to answer. An authorised reviewer may need protected access to personal, proprietary or security-sensitive material. Where governing law permits, redaction, severance, confidentiality undertakings, secure inspection, closed technical sessions and court-controlled expert access can reconcile these needs. None excuses failure to preserve records or prevents the adjudicator from examining indispensable evidence. If a party or vendor withholds necessary records under its control, the response depends on the forum and governing law. Production orders, reduced evidentiary weight, legally justified adverse inferences, exclusion, remand or interim relief may be appropriate. Sanctions must remain connected to burden, relevance and fairness.

Table 4.4 | Preservation and disclosure records

Record group	Preserve at source	Minimum party-facing disclosure	Controlled reviewer access	Question enabled
Authority and purpose	Enabling provision, approval, policy and intended consequence	Legal basis, criterion and responsible body	Full approval and delegation chain	Was power lawfully exercised?
Inputs and provenance	Case snapshot, source, match history, corrections and timestamps	Material attributed facts and their sources	Lineage, matching rules and source-quality evidence	Are these the right and accurate facts?

Record group	Preserve at source	Minimum party-facing disclosure	Controlled reviewer access	Question enabled
Model and configuration	Dataset / model / version identifiers, dependencies, features, threshold	Automation function, significant factors and material threshold in intelligible form	Model card, configuration, code/weights where necessary, reproducible environment	What produced this output?
Validation and error	Test design, population, subgroup results, calibration and limitations	Material known uncertainty or limitation	Full validation, drift and error analysis	Was use reliable in this context?
Output and human action	Score/class, explanation settings, user view, overrides and final act	Output's role, official's conclusion and reasons	Event trail, explanation reproducibility and intervention history	Did a person exercise judgement?
Security and incidents	Access logs, alerts, integrity evidence, changes and incident file	Material compromise affecting the case, subject to lawful limits	Vulnerability, forensic and containment records	Was evidence or operation compromised?
Contract and vendor	Specifications, audit clauses, subcontractors, updates, retention and exit	Relevant responsibility and complaint contact	Technical schedules, representations, audit and production rights	Who controlled evidence and risk?
Review and remedy	Notice, representations, correction, review and implementation	Forum, time limit, accessible route and outcome	Complete complaint and remedial record	Could error be effectively corrected?

Relevance authenticity and weight of machine-generated evidence

Pakistani evidence law accommodates electronic material while retaining conditions for its use. Article 46-A of the Qanun-e-Shahadat Order 1984 makes statements in electronic documents generated, received or recorded by an automated information system relevant when the system was working properly. Article 59 makes expert opinion relevant on science, electronic authenticity and integrity, and the functioning, specifications, programming and operation of information systems. Article 73, Explanation 3, prevents an automated output from losing primary-evidence status solely because it is electronic. In the absence of contrary evidence, it also presumes that the system was working at material times. Article 78-A governs proof where an alleged electronic signature or document is denied. Article 164 allows a court, according to the circumstances, to admit evidence or witnesses recorded through modern devices or techniques. The ETO

likewise prevents denial of recognition or admissibility merely because material is electronic (s. 3).

These provisions remove an objection to the medium. Relevance, authenticity, reliability and evidential weight remain distinct questions. A working system does not establish representative training data, a correct biometric identification, a calibrated risk score, a faithful summary or a valid legal inference. In Mian Khalid Pervaiz v State (2021 SCMR 522), the Supreme Court treated electronically generated defence material as capable of admission under Articles 46-A, 78-A and 164 and the ETO, while requiring disputed material to be proved according to law. Ishtiaq Ahmed Mirza v Federation of Pakistan (PLD 2019 SC 675) emphasised genuineness, accuracy, source, identification, safe custody, forensic examination where needed and production through competent witnesses for audio and video. These requirements become particularly relevant when cheap generation and manipulation make appearance an unreliable guide.

The court should first identify the proposition the output is offered to prove. Authorship, identity, file integrity, system operation, the meaning of a score and an element of liability each require different foundations. Authentication may need the native file, original device or account, metadata, creation and transfer history, hashes or signatures, access controls and chain of custody. A cryptographic hash can show that a preserved file has not changed since hashing. It does not show who created it, whether its contents are true or whether manipulation preceded capture. An authentic log can still omit context. A reproducible score may still come from a model invalid for the relevant population or purpose.

Evidence of error must address the disputed use. Aggregate accuracy is rarely sufficient. A court may need false-positive and false-negative rates, calibration, uncertainty, subgroup and language performance, validation populations, drift, thresholds and the exact model version. The relevant inquiry concerns performance on materially comparable records and conditions, rather than earlier success on a vendor dataset. Probability is not fact and does not decide the legal standard of proof. A risk score carries only the meaning supported by its model, target, time horizon and calibration. It is no probability of guilt unless the measure validly supports that meaning. In criminal proceedings, an output neither transfers the prosecution's burden nor dilutes proof beyond reasonable doubt. QSO Articles 117–120 allocate burdens for asserted and foundational facts. Article 122, on facts especially within a person's knowledge, should not casually be used to require an accused to disprove a vendor-controlled system.

Expert evidence should make machine-generated proof open to testing. It should not lend authority to material nobody can inspect. Experts should state their qualifications, instructions, data and version examined, methods, assumptions, validation, uncertainty, alternative explanations and limits, and remain open to cross-examination. Reading a vendor report alone cannot independently establish provenance or performance. Source

code is not always necessary either. The dispute may concern wrong inputs, a threshold, deployment conditions or human misuse. Disclosure should address the proposition in issue proportionately while allowing a real challenge. Meera Shafi v Ali Zafar (PLD 2023 SC 211) illustrates the wider principle: technology can assist the giving of evidence, but identity, voluntariness, examination and judicial supervision remain essential. A change of medium leaves the court responsible for fairness.

Synthetic media and generated documents

An allegation of synthetic audio, video or imagery cannot be resolved by choosing between visual intuition and a detector score. The court should test at least three possibilities: the item is authentic; it was generated or materially altered; or authentic material was deceptively edited, captioned or taken out of context. The proponent should preserve and, where required, produce the native file, capture and source records, metadata, transfer history, hashes and corroboration. A competent forensic examiner should identify the tool and version used, validation and error limits, and transformations that may affect detection. The opinion should distinguish a finding of manipulation or provenance from a mere absence of detected artefacts.

Watermarks, content credentials and deepfake detectors can fail. Provenance metadata may be omitted or stripped, watermarks may degrade, and detectors may perform poorly across generators, compression methods or changing distributions. NIST treats authentication, provenance tracking, labelling and detection as complementary techniques whose value depends on context (Chandra et al., 2024). No detected manipulation does not prove authorship, context or truth. Generated documents require similar care. Lawyers and judges must verify each authority against an authoritative report or official statutory text, check quotations and negative treatment, and keep factual statements within the record. Where relevant to a contested filing, decision or incident, material prompts, inputs, outputs and model/version details should be preserved. Governing records, confidentiality, privilege and litigation-hold duties continue to apply.

AI in courts and legal practice

Judicial AI needs the same status distinctions used elsewhere in this book. In Ishfaq Ahmed v Mushtaq Ahmed (PLD 2025 SC 582; 2025 SCP 112), the Supreme Court cautiously welcomed AI as an aid, identified risks associated with its use and called for institutional guidance. It conferred no adjudicative power on software. The judgment supports a careful distinction between assistance with judicial work and the judge's responsibility to decide. A proposal for guidance does not establish a completed institutional policy or the operation of every possible research, administrative or drafting system.

Evaluation should precede any claim that judicial AI improves the administration of justice. A bounded trial should identify the courts, participants, tasks, training, comparison process and period examined. Measures of speed or cases resolved need to be considered alongside accuracy, fairness, reason-giving, confidentiality and the ability to correct mistakes. Reported uptake alone does not establish safer adjudication. Nor can an aggregate result validate a particular prompt, draft or judgment. Institutions should therefore retain task-level evidence and keep judges able to reject the system's output.

Automation may help allocate and schedule cases under legally and institutionally approved criteria. Assignment rules, any required randomisation, conflicts, priorities, manual overrides and changes should be recorded and auditable. A vendor should have no ability to profile likely outcomes, steer cases towards a preferred bench or alter allocation unseen. Judicial leadership remains responsible for the allocation arrangements. Judges retain institutional independence in deciding cases (Constitution of the Islamic Republic of Pakistan, 1973, art. 175(3)).

Where rights or evidence depend on exact words, generated transcription and translation should be treated as drafts. The original audio or video and source-language record should be preserved. Speakers, timestamps, confidence and disputed passages should be identifiable, with competent human verification or certification as procedure requires. A translated summary must not silently replace testimony. Parties need a way to propose corrections, especially when accents, code-switching, Urdu or regional languages, disability-related speech, poor audio or specialised terminology increase error risk.

An authorised research or drafting system may retrieve, summarise or improve text. The judge must still read the underlying record, verify cases and legislation, decide the issues and adopt the reasons. Article 191 empowers the Federal Constitutional Court and Supreme Court to regulate their respective practice and procedure, and Article 202 empowers the High Courts, subject in each case to the Constitution and applicable law. An external platform's terms cannot exercise that power (Constitution of the Islamic Republic of Pakistan, 1973, arts. 191, 202). Independence also requires institutional control over confidential files, model access and updates. Open justice may require disclosure of a material automated role in decision-making, without ordinarily requiring publication of judicial deliberations or every research query. Disclosure should enable testing of a material automated influence or contested factual source, preserve what review requires and protect legitimate deliberative and security interests under law.

Machine-assisted ingestion, de-duplication, classification and retrieval can reduce the work of managing evidence. Original items, metadata and chain of custody must remain intact. Search rankings should not quietly determine legal relevance. Quality assurance should test whether the system systematically misses exculpatory, adverse or minority-language material. A human remains responsible for disclosure, privilege review

and the evidence put before the court. Consistently with Article 25 and Section 4.2, validation and monitoring should examine equal access and differing error rates across relevant languages, disabilities, genders, regions and other legally material groups. Efficiency cannot justify repeatedly obscuring a class of litigants or evidence.

Advocates retain their existing duties when using AI. The Pakistan Bar Council's canons require devotion and the advocate's utmost learning and ability within law (r. 156), impose duties to the court (r. 159), and prohibit intentional misquotation of testimony, documents and authorities or knowing reliance on invalid authority without disclosure (r. 161). Prosecutors must not suppress exculpatory facts (r. 163), and breach of the canons constitutes professional misconduct (r. 175-A) (Pakistan Legal Practitioners and Bar Councils Rules, 1976). Rule 146 protects confidential information in the context of conflicts of interest, while QSO Articles 9–12 govern professional communications and privilege. Although drafted before generative AI, these provisions leave lawyers responsible for accuracy, fidelity, candour and confidentiality. Before entering client information into a tool, counsel must know who receives it, whether it is retained or used for training, and whether professional secrecy and client instructions permit transfer. Staff, vendor and model output all require competent supervision and source verification.

Comparative guidance can help identify the questions. UNESCO's judicial guidelines emphasise assistance, meaningful human supervision, rights, transparency and institutional governance (Gutiérrez, 2025). Formal Opinion 512 of the American Bar Association Standing Committee on Ethics and Professional Responsibility addresses competence, confidentiality, communication, supervision, candour and reasonable fees in generative-AI use (American Bar Association Standing Committee on Ethics and Professional Responsibility, 2024). Neither governs Pakistani courts or advocates. Pakistani rule-makers, Bar Councils and judicial institutions can answer those questions under local law.

A legally usable output must pass through an identifiable process of authority, proof and human judgement. Notice must reveal the case to answer, and reasons must connect facts with law and consequences. Reconsideration needs real power to depart and correct. Preservation must allow both system behaviour and institutional action to be reconstructed. Courts should apply the governing law when admitting electronic or machine-generated material, then test its precise proposition, source, integrity, error and weight. Assistive tools may support judges and lawyers. Adjudication, advocacy, candour, confidentiality and the record remain human and institutional responsibilities.

4.4 Attribution of Responsibility, Forums, Standards of Review, and Remedies

An automated system cannot hold jurisdiction, answer a complaint, satisfy a judgement or restore a benefit. We must look to the people and institutions that supply data, set purposes, procure and configure the system, choose thresholds, secure it, interpret results, act and hear challenges. This does not mean that one party is liable for everything. Public law, contract, negligence, regulation, professional discipline, employment law and criminal law impose different duties through different forums. Several may apply to the same events. None should be invented merely because AI is involved.

The inquiry in this section connects legally recognised harm to the governing duty, the actor with control at the relevant stage, the supporting records, the competent forum and the available remedy. Review is strongest when those connections are documented before a dispute. Comparative scholarship likewise treats automated decision-making as a lifecycle involving institutions, people and technology, rather than a model alone (Cobbe et al., 2021). That approach is useful in Pakistan as a method of organising evidence. It creates no foreign cause of action.

Public-law authority, delegation, and retained responsibility

Legality is the first duty in public administration. Article 4 protects treatment in accordance with law, while Articles 8 and 10A and the context-specific guarantees considered in Section 4.2 constrain public power (Constitution of the Islamic Republic of Pakistan, 1973, arts. 4, 8, 10A). A model score, vendor recommendation, policy target or contract cannot supply a missing statutory power. An authority empowered to investigate fraud but not decide eligibility cannot use an anomaly detector to turn suspicion into disqualification. Where law entrusts clinical or licensing judgement to a professional, a software default cannot silently take its place. The enabling statute, valid delegated instrument, rules of business and authorised decision process must identify who may act, for which purpose, on what considerations and through which procedure.

Technical assistance and delegation of public power are different arrangements. Where law permits, an institution may contract for development, hosting, data cleaning, security monitoring or analytical support. The contract does not thereby authorise a supplier to exercise statutory discretion. Section 24 of the Digital Nation Pakistan Act 2025 illustrates the need to read the particular statute. It permits the Pakistan Digital Authority to delegate a specified power by Gazette notification, within stated limits and a stated period, only to its Chairperson, a Member or an officer. It does not list a vendor. This provision is no universal delegation code, but it shows why outsourcing cannot simply be assumed to transfer decision-making authority.

The public body must retain control appropriate to its function. At minimum, it should approve the purpose and material criteria, know which data and model versions operate, and secure relevant evidence. It should state when outputs may be followed, questioned or disregarded, identify the authorised official, record final reasons, and retain powers of reconsideration and correction. Relief must be possible without supplier consent. A signature or click means little where the official lacks time, access, competence or authority to depart (Binns, 2022). Meaningful official review does not, in turn, excuse an inaccurate vendor warranty or negligent professional reliance. Public and private duties can apply within the same process.

At federal level, attribution of executive action must respect Articles 90 and 99 and the Rules of Business. Mustafa Impex, Karachi v Government of Pakistan (PLD 2016 SC 808) establishes that the Federal Government is the constitutionally defined collective entity. Authentication cannot replace the authority and decision process required by law (Constitution of the Islamic Republic of Pakistan, 1973, arts. 90, 99). The same distinction applies to an automated programme: a technically authenticated output is not necessarily a lawfully authorised act. Approval of a vendor system by an IT unit cannot validate an otherwise incompetent decision. Nor may an official act in the Federal Government's name where law and the Rules of Business require a different approval.

Once authority is established, we must examine how it was exercised. The governing law and pleaded right determine the questions. Did the institution follow mandatory procedure, notify or hear the affected person, apply the statutory criterion and consider material contrary evidence? Did it rely on an irrelevant proxy, mistake probability for proof, provide intelligible reasons, maintain equality or act proportionately where required? Was the output adopted mechanically? An accurate model can contribute to an unlawful decision by automating the wrong criterion. A decision can also fail because the source record was wrong, even when the model processed it as designed. Public-law responsibility concerns the administrative act and procedure. It does not depend on proving a malfunction in source code alone.

A private supplier may be relevant to constitutional proceedings, but its label does not settle the matter. Article 199 requires a functional inquiry into whether the entity was entrusted with State functions involving public or sovereign power, taking substantial government control and State funding into account. These considerations do not bring every private dispute within constitutional review (Salahuddin and 2 others v Frontier Sugar Mills and Distillery Ltd., Takht Bhai and 10 others, PLD 1975 SC 244, p. 257). A software licensor is therefore no constitutional authority merely because it licenses software. Where procedure allows, it may still be joined, required through the public respondent or applicable process to produce evidence, or face separate contractual or civil liability. Crawford and Schultz (2019) describe the gap created when the State disclaims technical knowledge and the vendor denies public responsibility. Attribution

and evidence access should close that practical gap under Pakistani law while preserving the parties' distinct legal identities.

Contractual, civil, consumer, employment, professional, and criminal responsibility

Contracts can allocate duties directly among purchasers, developers, integrators, cloud providers, data suppliers and cybersecurity contractors. They can warrant provenance, allocate security, notification and correction duties, preserve versions and incident records, provide audit access, control updates and subcontractors, and set indemnity, transition and termination rights. Section 73 of the Contract Act 1872 permits compensation for loss naturally arising from breach or known at formation as likely. It excludes remote and indirect loss, with mitigation remaining relevant. Section 74 limits recovery under a stipulated sum or penalty to reasonable compensation within the named amount. A precise, testable warranty is therefore more useful than a general promise of an ethical or accurate product.

A contract ordinarily binds its parties. Without a recognised exception or independent statutory or public-law duty, a non-party generally cannot enforce a public procurement contract simply because its terms benefit the public (Talaat Inayatullah Khan and another v Dr Anis Ahmad Sheikh, PLD 2015 Sindh 134, pp. 137–138, para. 10; Messrs Mastersons through its Partner v Messrs Ebrahim Enterprises and another, 1988 CLC 1381, p. 1388; Khokhar, 2024). A claimant must identify an applicable public, statutory, contractual, consumer, tortious, employment or other right. Purchaser and supplier may allocate specified losses through caps and indemnities. They cannot authorise coercive public action, extinguish statutory protection, limit the Auditor-General's lawful access, prevent a competent court or regulator from obtaining material, or excuse reconsideration by the public body. A service credit does not fully remedy discriminatory denial, unlawful disclosure or corrupted records.

Negligence may provide a civil route, although Pakistan has no general AI-specific tort or automatic strict liability for algorithmic harm. The applicable claim requires a recognised duty, breach of the governing standard, factual and legally attributable causation, and recoverable damage. In National Logistic Cell v Irfan Khan, the Supreme Court applied reasonable care, required proof of breach and resulting damage, and considered negligence by actors whose conduct combined to cause loss (2015 SCMR 1406, paras. 15, 18–20, 22). Control helps identify possible precautions in an AI supply chain. A data provider may address provenance, a developer architecture and testing, an integrator configuration, and a cybersecurity provider monitoring and escalation. Deployers control purpose, thresholds and human workflows; professionals control their interpretation of results. Each duty must still arise from the pleaded relationship and

governing law. Comparative work can identify who might prevent or insure a risk, but creates no Pakistani cause of action (Buiten et al., 2023).

Provincial consumer statutes may impose more specific duties for purchased products or services. The Punjab Consumer Protection Act 2005 addresses faulty or defective services, product responsibility, disclosure and redress within its provincial framework. Territorial regimes differ, so definitions, notice, limitation, jurisdiction, defences and available orders must be checked. A person denied a bank loan, overcharged by a platform or misled about an automated service may also have a sectoral route in banking, competition, telecommunications or securities law. Section 10 of the Competition Act 2010, for example, gives the Competition Commission jurisdiction over deceptive marketing as market conduct and public enforcement. That jurisdiction should not be assumed to replace an individual damages action or Banking Mohtasib complaint.

Employment duties also depend on the legal relationship. Recruitment rankings, scheduling, productivity monitoring, discipline and dismissal operate within applicable contracts, standing orders, industrial-relations legislation, public-service and workplace-harassment law, constitutional limits on State employment, and professional or sector rules. The Industrial Relations Act 2012 applies within its federal reach. After devolution, provincial laws govern many other workplaces. A covered public servant's terms-and-conditions dispute may belong before departmental authorities and a service tribunal rather than in an ordinary civil or constitutional merits appeal. A private worker may use internal procedures, collective machinery, a labour court or the National Industrial Relations Commission only where legislation gives jurisdiction. Pakistan has no independent claim named algorithmic employment unfairness. The practice must be connected to a recognised contractual, statutory, constitutional, discriminatory or tortious wrong.

Professionals retain personal responsibility when software assists them. A doctor using clinical support, an advocate submitting generated authority, an accountant approving an automated compliance report and an engineer accepting a risk classification remain subject to their governing legislation and professional standards. Under the Pakistan Medical and Dental Council Act 2022, the Pakistan Medical and Dental Council may investigate within its jurisdiction. In Punjab, the Healthcare Commission may investigate conduct or service quality within its jurisdiction under the Punjab Healthcare Commission Act 2010. Bar Councils may discipline advocates under the Legal Practitioners and Bar Councils Act 1973. Discipline examines competence, candour, confidentiality, supervision and care. Discipline does not necessarily compensate an injured person, reverse the underlying administrative act or determine a vendor's breach. Those outcomes may need parallel proceedings.

Cybersecurity and criminal routes have narrower conditions. Unauthorised access, copying or transmission, interference, identity misuse, electronic fraud, malicious code

and related conduct may fall under PECA 2016, as amended, or general criminal law if every element is proved. A poisoned dataset, extracted model, leaked inference or fabricated record is not criminal merely because it causes technical harm. Authorisation, conduct, mental element, jurisdiction and evidence remain decisive. Investigation may preserve evidence and punish established wrongdoing. It does not by itself correct eligibility records, restore services, compensate everyone affected or redesign procurement. Those remedies require separate action.

Causation, proof, and vendor-controlled evidence

Where several actors contribute, causation should be reconstructed carefully. A chronology should follow the definition of need, procurement and acceptance, source records, training and labelling, model and threshold approval, integration, security controls and incidents. It should then trace the case input and output, human review, final action, downstream sharing, complaint and remediation. At each stage, identify control, duty, the contemporaneous version and the counterfactual. Would the result have occurred without the wrong record, classification, threshold, interface default, compromise, professional reliance or official act? Several failures may have contributed. The inquiry must also ask whether the loss fell within the duty's scope and was sufficiently foreseeable or proximate.

The legal proceeding determines how these questions are answered. Constitutional review may succeed because an adverse public act lacked authority or fair procedure, without identifying the precise mathematical origin of an error or quantifying civil damages. A purchaser claiming in contract must connect breach to recoverable contractual loss. Negligence requires its own elements of causation and damage. Regulatory proceedings follow the regulator's statute, evidence powers and prescribed standard. Criminal guilt requires proof beyond reasonable doubt; civil and administrative proceedings follow their respective laws.

The Qanun-e-Shahadat Order 1984 sets general rules for proof. Articles 117–119 place burdens on the party asserting facts or the party who would fail without evidence. Article 122 places the burden of a fact especially within a person's knowledge on that person. Article 129(g) permits, without compelling, an inference that available evidence withheld by a person would have been unfavourable to them. These provisions may matter where only a vendor or deployer knows the operative version, whether a warning was suppressed or why logs disappeared. They do not automatically reverse the burden in every opaque-system dispute. Special knowledge, withholding, preservation duties and any proposed inference must be related to the pleaded issues and surrounding facts.

Civil procedure can provide practical access. Section 9 of the Code of Civil Procedure 1908 preserves jurisdiction over suits of a civil nature unless cognisance is barred. Order XI permits interrogatories and court-ordered discovery, production and inspection of

documents in a party's possession or power where necessary for fair disposal or saving costs. Privilege and relevance remain grounds for objection, and non-compliance may have procedural consequences. The phrase 'possession or power' matters when a public body can contractually require its supplier to produce logs. Procurement should make that control express. The institution should not discover after harm that it has left the vendor with exclusive control of the evidence.

As Section 4.3 explains, electronic form and expert-evidence provisions do not settle authenticity or reliability. For causation, controlled access is essential. Confidentiality, cybersecurity, personal data, privilege and trade secrets may justify redaction, undertakings, restricted expert inspection or other lawful safeguards. They do not justify concealment from every competent reviewer. Courts, regulators and experts may need inputs, outputs, lineage, model and configuration versions, validation, thresholds, changes, overrides and incident records. Source code alone may not describe the deployed system. Integrity, reproducibility, error evidence and expert method continue to affect weight (Rasool & Rasool, 2022).

Missing records can have several legal consequences. They may support an adverse inference, prevent an authority from showing lawful reasons, breach a records duty, weaken evidential weight, justify regulatory action or require preservation and independent examination. They do not invariably prove discrimination, negligence, causation or damages. The response should require the party with control to account for the record, protect legitimate secrets and draw only justified inferences. Uncertainty caused by poor governance must remain distinct from affirmative proof of the ultimate allegation. There is no need to invent a general AI presumption.

Forums and standards of review

The duty and requested remedy determine the route. As Section 4.3 explains, internal reconsideration is often the quickest way to correct records, receive evidence, pause adverse action or decide the merits afresh. It requires a lawful reviewer with the operative record, suitable competence and power to depart and grant relief. Requirements to use another remedy first depend on the statute. Urgency, futility, jurisdictional defects and fundamental-rights claims may affect how they apply.

Sectoral review provides expertise within defined limits. Telecommunications complaints may engage PTA's licensing, information, complaint, direction and enforcement powers under the Pakistan Telecommunication (Re-organization) Act 1996. Where applicable, the route continues to the Telecommunication Appellate Tribunal and statutory appeal to the Supreme Court under the 2024 Act. Banking complaints may begin with the bank and proceed within sections 82A–82G of the Banking Companies Ordinance 1962 to the Banking Mohtasib for specified malpractice, arbitrary or discriminatory action, violations, delay or maladministration. Subject to

statutory exclusions, the Mohtasib may direct reconsideration, modification or cancellation, reasonable compensation and operational improvement. Sections 13–14 of the Federal Ombudsmen Institutional Reforms Act 2013 allow review by the Mohtasib and representation to the President, each sought within thirty days. The State Bank Governor is not the appellate forum. Section 82F of the 1962 Ordinance permits requests for relevant documents and an adverse inference if a bank refuses, rather than direct compulsion. SECP decisions follow the internal review, Appellate Bench and court routes in sections 32B–34 of the Securities and Exchange Commission of Pakistan Act 1997. Competition orders proceed through the Appellate Bench, Competition Appellate Tribunal and Supreme Court under sections 41–44 of the Competition Act 2010. Tax, customs, health, education, labour, consumer and professional disputes each require their own enabling and territorial route. A regulator's existence in one field creates no general AI jurisdiction.

Ombuds, rights, information and audit bodies serve different purposes. The Wafaqi Mohtasib investigates maladministration by covered federal agencies under the Establishment of the Office of Wafaqi Mohtasib (Ombudsman) Order 1983, subject to exclusions, alternative-remedy provisions, review and representation under the Federal Ombudsmen Institutional Reforms Act 2013. Provincial Ombudsmen act under their respective statutes. They may expose delay, opacity, mechanical reliance and failures to correct, without providing universal model certification. The National Commission for Human Rights may investigate rights violations or related public-servant negligence, use statutory civil-court powers to compel evidence, recommend prosecution or other action and immediate interim relief, and report on systemic problems under the National Commission for Human Rights Act 2012 (ss. 9, 13, 18). Its recommendations are neither licensing sanctions nor ordinary merits appeals. Information commissions enforce access within the coverage and exemptions of federal or provincial law; disclosure alone does not remake the challenged decision. Under Articles 168–171 and the Auditor-General's (Functions, Powers and Terms and Conditions of Service) Ordinance 2001, the Auditor-General examines legality, expenditure, performance, controls and relevant records and reports for legislative scrutiny. Audit ordinarily restores no individual benefit and awards no tort damages.

Specialised tribunals and commissions have only the jurisdiction conferred on them. Article 212 and the Service Tribunals Act 1973 give service tribunals exclusive jurisdiction over covered civil servants' terms and conditions, including discipline, subject to departmental recourse and statutory exclusions. Labour courts, revenue and tax appellate bodies, consumer courts, healthcare commissions, professional councils, procurement grievance bodies and other tribunals differ in standing, time limits, evidence, interim powers and appeal. Partial mandates should not leave a complainant moving endlessly between institutions. The deployer should identify a lead route,

preserve limitation periods, transmit evidence lawfully and explain what relief each body can and cannot provide.

Constitutional and ordinary courts complete the institutional account in Chapter 1. Under Article 199, a High Court may, on the applications and standing terms specified in its several limbs and within its conditions, compel or restrain public action, declare an act without lawful authority, address unlawful custody and enforce fundamental rights. Article 202A(3) reserves Article 199 jurisdiction to Constitutional Benches where Article 202A has been brought into force under clause (7). Article 175E gives the Federal Constitutional Court (FCC) distinct intergovernmental, fundamental-rights and call-up jurisdictions, while Article 175F governs constitutional appeals. Intergovernmental relief is declaratory. Fundamental-rights jurisdiction is application-based and requires the FCC to consider that a question of public importance relating to enforcement of fundamental rights is involved. Subject to Article 175F, the Supreme Court retains Article 185 appellate jurisdiction and statutory appeals in non-constitutional fields (Constitution of the Islamic Republic of Pakistan, 1973, arts. 175E–175F, 185, 199, 202A). As Chapter 1 explains, amended Article 189 makes an FCC decision binding on every other court, including the Supreme Court, insofar as it decides a legal question or states a legal principle. A Supreme Court decision on such a matter binds every other court except the FCC. These provisions allocate forums. They appoint neither court as a standing AI regulator.

The standards of review differ. A statutory appeal may permit correction of law, fact or discretion to the extent specified by its Act. Merits reconsideration may substitute a fresh authorised decision. Constitutional review principally examines authority, procedure, rights and legally reviewable abuse; it does not ordinarily rerun technical administration as a first-instance merits tribunal. Civil courts determine recognised private rights and may award damages or injunctions within jurisdiction. Order XXXIX of the Code of Civil Procedure 1908 provides temporary injunctive machinery in covered cases. Criminal courts determine charged offences through criminal process. The FCC's power under Article 175E(5) to call for a record is no standing authority to audit every algorithm. Continuing judicial or regulatory supervision should likewise rest on lawful jurisdiction and demonstrated systemic non-compliance, with clear milestones, reporting, independent expertise where permitted and an exit condition.

Matching remedy to cause, urgency, scale, and recurrence

An effective remedy addresses the immediate injury and the process likely to repeat it. Urgency concerns threats to liberty, subsistence, treatment, housing, education, employment, account access, evidence or security. Reversibility asks whether later payment or reconsideration can repair the loss. Scale concerns whether one record or an entire population is affected. Recurrence asks whether the cause lies in an isolated input

error, shared database, threshold, model behaviour, unauthorised purpose or institutional workflow. These differences should guide the response.

For an urgent but reversible individual error, preservation, a temporary pause, assisted record correction and prompt human reconsideration may be proportionate. If the loss would become irreparable, interim continuation of a benefit or service, restraint of disclosure or enforcement, or another competent protective order may be required. Final relief may include intelligible reasons, source correction and propagation, a fresh lawful decision, restored benefit or status, legally permitted deletion or restriction, a declaration or an injunction. Compensation may be available under an applicable statutory, contractual, tortious, ombuds or constitutional remedial basis. Automation alone does not establish entitlement to it. Jurisdiction, causation, loss, immunities, limitation and the governing cause of action still control.

Recurring harm or harm across a population may require wider action. Within its powers, a competent body may order preservation and independent audit, notice to affected people, suspension of a use, data feed or automated consequence, revised criteria or thresholds, record cleansing and revalidation across groups. Other responses may include retraining on lawfully obtained data, rollback to a validated version, more human staff, incident notification, vendor cure, indemnity or termination, procurement revision or re-procurement, professional or disciplinary action, policy or rule changes, continuing reports or decommissioning. A vague direction to debias the AI cannot guide compliance. Relief should identify the institution, defect, evidence required, corrective action, deadline, verification method, treatment of pending cases and route if compliance fails.

Deletion and decommissioning need careful sequencing. Immediate deletion may destroy evidence, breach an archival or litigation hold, or prevent a correction from reaching linked systems. Indefinite retention may prolong surveillance and breach risks. A remedy should distinguish operational use, restricted preservation, downstream correction, lawful retention and eventual verified deletion. Terminating a vendor without a tested manual or replacement route can also interrupt essential services. Decommissioning must provide continuity, export records and configurations, revoke access, dispose of models and credentials, preserve unresolved claims and keep the transition under public responsibility.

The matrix identifies likely primary routes. Each remains subject to the enabling law, territorial competence, standing, limitation, exclusions and facts of the case.

Table 4.5 | Attribution forums and matched remedies

Responsible actor or chain point	Principal duty or possible breach	Minimum proof	Primary forum or scrutiny route	Matched individual and systemic remedies
Deploying public institution and authorised official	Lawful power; correct purpose and criteria; fair procedure; reasons; real exercise of discretion; record and remedy retention	Enabling law; approval; input/output and version; threshold; human action; reasons; complaint record	Internal / statutory reconsideration; sector appeal; ombuds; High Court (Constitutional Bench where Article 202A is in force); FCC where Article 175E applies	Pause; reasons; correction; fresh lawful decision; restoration; declaration; systemic rule/workflow change
Data custodian or data provider	Lawful supply; provenance; accuracy and correction; security; stated contractual or sector duties	Source and change history; sharing authority; match; access log; correction propagation	Custodian complaint; deploying body; information commission where applicable; regulator; civil/contract route	Correct and notify downstream users; restrict use; compensate where legally available; improve quality and sharing controls
Developer, vendor, integrator, cloud or cybersecurity provider	Specification, warranty, reasonable care, secure integration, monitoring, disclosure, change and incident duties	Contract; representations; design/test record; model / configuration version; vulnerabilities; incident and access logs	Contract dispute / arbitration as agreed; civil court; procuring agency/PPRA route for pre-contract procurement acts; sector regulator; criminal route if offence proved	Cure; evidence production; damages / indemnity; security repair; rollback; termination; transition; debarment or sanction where law permits
Professional user or employer	Professional competence and supervision; independent judgement; employment and workplace duties; no mechanical reliance	Applicable standard; warning / interface; time and authority; intervention; employment or clinical record	Employer / internal route; professional council; healthcare commission; labour/National Industrial Relations Commission (NIRC) or service tribunal; civil court	Reconsideration; correction; discipline; training / supervision; reinstatement or compensation where authorised; workflow redesign
Bank, telecom licensee, platform or other regulated provider	Sector statute, licence, directions, consumer protection, non-discrimination, complaint and security duties	Transaction and account record; licence / direction; model and threshold; notice; customer complaint; impact	Banking Mohtasib/SBP; PTA and Telecommunication Appellate Tribunal where applicable; SECP; Competition Commission of Pakistan (CCP); provincial consumer forum; statutory appeal	Reverse charge/action; reconsider; compensation where empowered; direction, penalty, licence action; class notice and control change

Responsible actor or chain point	Principal duty or possible breach	Minimum proof	Primary forum or scrutiny route	Matched individual and systemic remedies
Institution-wide or repeated failure	Governance, procurement, audit, rights, record, monitoring, or coordination failure	Pattern of cases; subgroup and reversal data; audit trail; common model / data / version; prior notice and failed cure	Regulator; National Commission for Human Rights (NCHR); Ombuds; Auditor-General / Public Accounts Committee (PAC); constitutional court within jurisdiction; legislative committee	Independent audit; suspension; affected-class review; procurement/rule reform; continuing supervision; re-procurement or decommissioning

Pakistan need not assign legal responsibility to a machine or assume that one new institution or claim will resolve every automated harm. Duties must remain with actors able to perform them. Evidence must remain accessible across organisational boundaries, and claimants must reach forums capable of granting relief. Where defects recur, correction must extend beyond one file. Accountability depends on matching the response to the wrong and preventing contractual secrecy, divided roles and missing logs from making an existing right impossible to enforce in practice.

Chapter 4 Key Arguments

Evidence comes first. AI use in Pakistan must be described through documented status and consequences, rather than novelty, promotion or technical possibility. Announced, proposed, procured, piloted, operational, suspended, discontinued and unverified each describe a different proposition and require different evidence. The institution, territory, population, function, version and date must remain clear. A tender shows intended acquisition, without proving production performance. A research prototype shows feasibility in its study setting, without proving government adoption. A policy commitment does not establish an operating system. Where operation is undocumented, uncertainty is the proper finding. The same discipline prevents ordinary digitisation, databases, ATMs, rules engines and e-government services from being casually renamed AI.

Rights analysis follows the effects of the system and the legal character of the act. Collection, linkage, surveillance, matching, ranking, profiling and inference may affect dignity, security of person, privacy, fair process, information access, equality, reputation, livelihood, property, health, education and public services even without full automation. Lawful collection gives no unlimited authority to create and use risk labels. The actual use requires examination of authority, legitimate purpose, necessity, proportionality where applicable, security, retention and unequal effects. Equality analysis must address direct classifications, neutral rules with unequal burdens, proxies, representation and

measurement gaps, comparators and the unequal costs of error. Relevant differences may concern gender, disability, language, region, minority status, income, migration, documentation, rural residence and technological access. When linked institutions repeat one error, they create a systemic problem.

Review requires an intelligible institutional record. Notice should identify the act, legal basis, material data, automated role, significant factors, responsible body, human-review route, deadline and means of correcting source records. Reconsideration needs an authorised reviewer with relevant data and lifecycle records, time, competence and practical power to depart and restore the affected interest. Opacity, confidentiality, intellectual property and cybersecurity may justify proportionate protection. They cannot erase procedural fairness. Depending on the forum, review may require source and input records, provenance, dataset and model versions, validation and error rates, thresholds, logs, incidents, human interventions, policies, contracts and final reasons. Feature attribution or a generated explanation cannot supply, after the event, a lawful rationale that the official never had.

Legal persons and institutions bear responsibility. A deploying public body retains its function, the authorised official answers for the legal act, and professional users remain subject to competence and judgement standards. Developers, data providers, vendors, integrators, cybersecurity providers, employers, platforms and regulated firms may also owe distinct contractual, statutory, regulatory, professional, consumer, labour, civil or criminal duties according to conduct and control. Causation should be traced from source records and procurement through training, configuration, security, output, reliance and final action. Missing records or vendor control should prompt preservation, disclosure, expert assistance and legally available evidential consequences. Neither should create automatic immunity.

The wrong determines the forum and remedy. Internal review may reconsider merits; regulators may investigate regulated conduct; ombuds institutions may address maladministration; information commissions may order access; and auditors may examine controls or expenditure. Rights commissions may investigate and recommend, professional bodies may discipline, tribunals may exercise their statutory jurisdiction, and constitutional, civil and criminal courts apply their respective powers and standards. None provides a universal AI forum. Individual relief may require notice, corrected records, alternative access, reconsideration, restored benefits, restraint, legally available compensation or preserved evidence. Recurring harm may also call for independent audit, suspension, revised models or thresholds, retraining, contract enforcement, procurement reform, downstream correction, changed policies or rules, incident notification, continuing supervision or decommissioning. The remedy should correct the affected decision and, where recurrence is established, the data, model, contract, workflow or institutional practice that continues to produce the harm.

CHAPTER 05

Regulatory Coordination and Lifecycle Oversight in a Federal Legal Order

Introduction

The preceding chapters establish that artificial intelligence cannot be governed as a freestanding technical object. Chapter 1 located authority in Pakistan's Constitution, legislation, delegated instruments, institutional mandates and federal distribution of competence. Chapter 2 showed that records, identity systems, interoperability, procurement and administrative capacity shape what an automated system can know and what an institution can responsibly do. Chapter 3 added the neural-network lifecycle, cybersecurity threats, validation, explanation and auditability. Chapter 4 then followed consequential outputs into public administration and regulated sectors, connecting them to dignity, privacy, equality, fair procedure, evidence, responsibility, review and remedy. Chapter 5 converts those findings into an implementable governance architecture.

The central design problem is neither whether Pakistan should regulate AI nor whether one institutional model is universally superior. It is how common safeguards can operate across a federation in which constitutional and statutory powers, records, professional expertise, service delivery and remedies are distributed among federal, provincial, sectoral and local institutions. Capacity is distributed unevenly as well. A highly centralised body might concentrate scarce expertise and improve consistency, yet lack lawful jurisdiction or operational knowledge in provincial health, education, policing or municipal services. Purely sectoral governance can preserve expertise and lawful mandates, yet leave common procurement, security and evidentiary failures unaddressed. The chapter therefore develops a coordinated federal-sectoral model: shared vocabulary, minimum lifecycle controls and information exchange, combined with responsibility retained by the institution legally empowered to authorise, operate, supervise or review the use.

The starting legal position requires precision. The *National Artificial Intelligence Policy 2025* was an approved executive policy. It supports registers, impact assessment,

human oversight, audits, accountability, redress, sandboxes and federal–provincial coordination, but it is not an omnibus AI statute and cannot itself enlarge a regulator's jurisdiction or authorise coercive action (Ministry of Information Technology and Telecommunication [MoITT], 2025). The *Digital Nation Pakistan Act 2025* and existing sector statutes supply particular enacted mandates within their own scope (*Digital Nation Pakistan Act*, 2025). Comparative materials from UNESCO, the OECD, NIST and the European Union provide comparative design concepts, not Pakistani law.

Institutional fit consequently precedes institutional novelty. Common national functions may include a consequence taxonomy, interoperable registers, documentation and incident formats, model contractual clauses, shared technical capacity, research and consolidated reporting. Authorisation, licensing, professional supervision, service delivery, inspection, complaint handling and remedies should remain with competent departments, regulators, provinces, local bodies, ombuds institutions, audit bodies, commissions, tribunals and courts. Coordination cannot convert policy into law, and outsourcing cannot transfer official responsibility to a vendor. Provincial participation must be embedded in standard-setting, funding, implementation support and review where national systems rely upon provincial records or operate in devolved fields.

The chapter moves through the lifecycle. Section 5.1 creates functional and consequence-based scope. Section 5.2 allocates responsibilities through a federal-sectoral matrix. Section 5.3 makes registers, integrated assessment, authorisation and procurement conditions operate before acquisition, pilot or deployment. Section 5.4 joins monitoring and material-change review to incident governance, transparency, participation and graduated redress. Section 5.5 sequences reform, readiness, capacity, cost and periodic evaluation. Together, these controls make deployment conditional upon evidence and continued operation conditional upon performance.

The objective is not a maximal catalogue of formal duties. A rule that cannot be staffed, funded, evidenced or enforced may produce ceremonial compliance while increasing institutional dependence on suppliers. The proposed model therefore matches controls to consequence, specifies who must act, preserves lawful review routes, and treats suspension, correction and decommissioning as normal governance outcomes. Its measure of success is not the number of AI strategies or authorities created, but whether Pakistan can identify a consequential system, establish authority before use, reconstruct what occurred, respond to failure, provide an accessible remedy and discontinue a system that no longer meets its lawful purpose.

5.1 Regulatory Scope and Consequence-Based Classification

Pakistan needs a regulatory scope that can survive changes in model architecture and product vocabulary. A rule confined to "machine learning", "generative AI" or systems marketed as "autonomous" would invite obsolescence and strategic relabelling. One covering every formula, search function and digital workflow would exhaust scarce supervisory capacity while obscuring uses that alter liberty, entitlement, livelihood or access to an essential service. The appropriate unit is therefore the **system as used**: its function, institutional setting, exposed persons and interests, and resulting action. This approach is technology-neutral, but not consequence-neutral.

That proposition is a design recommendation, not a description of an existing omnibus regime. Pakistan had an approved *National Artificial Intelligence Policy 2025*, but no identified generally applicable AI statute establishing enforceable national risk tiers. The Policy calls for human oversight in critical or high-risk settings, registration of public-sector AI, lifecycle evaluation and impact assessment of high-impact systems, audits, accountability, redress and sandboxes. It does not define "critical", "high-risk" or "high-impact", and its call for a future legal framework confirms that policy commitments are not already enacted prohibitions or permissions (Ministry of Information Technology and Telecommunication [MoITT], 2025). The protocol below can guide procurement, administrative directions, sector rules and future legislation only through a body acting within lawful competence; it cannot create coercive power or validate an otherwise unlawful use.

Functional scope: governing what a system does

The gateway should cover a machine-based inferential system, and an adjacent automated decision system, when it performs a materially relevant function. These functions include **prediction** of an unknown or future condition; **classification** of a person, record, object or event; **ranking** of cases, candidates or options; **recommendation** of an outcome; **matching** of identities, records, characteristics or opportunities; **generation** of text, images, audio, code, summaries or purported reasons; **anomaly detection**; and **automated action** in a physical or digital environment. The OECD's classification framework usefully separates the context, data and input, model, and task and output dimensions, and stresses that an identical capability may present different implications in different settings (Organisation for Economic Co-operation and Development [OECD], 2022). It is a comparative analytical resource, not Pakistani law.

Functional coverage prevents two errors. A deterministic eligibility rule, watchlist match or compliance score should not escape because its supplier denies that it is AI: if it allocates a benefit, directs inspection or creates a practical presumption, its effects resemble those of a learned classifier. Nor should an advanced model itself attract the

highest tier. Translation used for an internal draft differs from unchecked translation of testimony; device-unlocking facial matching differs from a police watchlist search; and an informational chatbot differs from one that silently disposes of complaints. Classification attaches to each intended and reasonably foreseeable use, not the model in the abstract.

The protocol must also disaggregate a chain of functions. One application may match an identity, classify a file, score risk, rank it, recommend action and generate notice. Each output states a different evidential proposition and may need a different control. An anomaly may invite inquiry but not prove fraud; generated reasons cannot substitute for application of mind; clinical advice cannot silently displace professional judgement. The record should identify where output first influences a person or public function and every downstream action it can trigger. Otherwise, an institution may call a model "advisory" although interface, workload or policy makes acceptance routine.

Meaningful human control is a factual and institutional question. A reviewer needs lawful authority, competence, time, access to the record and system limitations, a workable ability to override, and an obligation to record actual reasons. A signature does not reduce consequence where these conditions are absent. Human review also cannot cure unlawful data, an impermissible criterion, population-wide surveillance or an exposed vulnerability. Comparative Article 14 of the EU Artificial Intelligence Act usefully treats understanding limitations, recognising automation bias, interpreting outputs, overriding and stopping a system as effective oversight; it does not apply in Pakistan by its own force (European Parliament & Council of the European Union, 2024, art. 14).

From consequences to differentiated controls

Classification should begin with the protected or regulated interest, not a vendor's probability of technical failure. In public administration the constitutional baseline includes treatment in accordance with law, security of person, fair trial and due process, dignity and privacy of the home, access to information and equality, according to the facts and applicable field (*Constitution of the Islamic Republic of Pakistan*, 1973, arts. 4, 9, 10A, 14, 19A, 25). Sector legislation, professional duties, licences, procurement rules and provincial law may add more specific standards. Effects upon investigation, detention, physical safety, legal status, public benefits, healthcare, education, employment, housing, credit, insurance or an essential utility should therefore carry greater weight than inconvenience in an optional, easily corrected service.

An arithmetic score should inform but not decide the tier: an average can conceal a disqualifying feature, give false precision to incomparable harms and offset a severe rights risk against commercial benefit. The protocol should combine **presumptive floors**, contextual modifiers and recorded judgement. A use that determines liberty, bodily

intervention, legal status, basic public services or serious investigation should ordinarily be no lower than prior assessment. Large scale, persistent biometric or health data, unavoidable exposure, cross-database linkage, difficult correction and repeated downstream use move it upwards. Genuine advice, ready reversibility and accessible reconsideration may reduce decisional risk only after evidence establishes them.

Six regulatory responses form a graduated but combinable control set. Ordinary and enhanced oversight describe baseline intensity; prior assessment adds pre-deployment proof; conditional authorisation and restrictions may operate together; prohibition is the terminal disposition. The protocol should therefore record both consequence level and the controls attached to the particular use:

Ordinary sector oversight is appropriate where the system performs a low-consequence assistive or administrative function, uses data lawfully handled in the ordinary course, does not materially structure a decision about a person, and can be corrected without significant loss. Existing records, procurement, cybersecurity, professional supervision and complaint rules remain applicable; "ordinary" does not mean ungoverned.

Enhanced documentation and monitoring applies where a system handles sensitive information, influences workflow or resource allocation, operates at appreciable scale, or could cause material but normally reversible harm. It requires an internal register, named ownership, documented purpose and data provenance, baseline and subgroup performance where relevant, security controls, logging, user training, monitoring and an escalation route.

Prior assessment is the default for a high-consequence use. Before acquisition, pilot or deployment, the competent institution should establish legal authority and necessity; assess rights, equality, data, security, procurement and institutional capacity; validate the system for the intended populations, languages and conditions; design meaningful human control, notice, correction and appeal; preserve evidence; identify the responsible official and reviewer; and approve a funded fallback and exit plan. NIST's voluntary AI Risk Management Framework similarly treats context mapping, measurement, governance and management as connected lifecycle functions, and warns that an inability to measure risk does not prove low risk (Tabassi, 2023).

Conditional authorisation is appropriate when a potentially valuable use retains material uncertainty that can be investigated within a bounded deployment. Conditions should fix the purpose, place, duration, case volume, eligible users, affected population, permissible consequence, performance and incident thresholds, independent evaluation, complaint access and automatic expiry. Continuation requires affirmative reauthorisation; a successful demonstration of technical performance does not establish legal authority or distributive acceptability.

Restricted use is required where only a narrower function or setting can be justified. Restrictions may confine use to specified officials, prohibit sole reliance, require independent corroboration, exclude particular populations or data, prevent downstream sharing, forbid real-time action, or permit research while barring operational deployment. The source and enforcer of each restriction must be identified. A classification form cannot supply a warrant, delegation or statutory criterion that the law withholds.

Prohibition is appropriate where a proposed public use lacks enabling authority or legitimate purpose; where a private or regulated use lacks any required lawful basis, licence or permission; where the use necessarily defeats an applicable constitutional, statutory or regulatory requirement; or where it presents severe, substantially irreversible residual harm that controls cannot acceptably reduce. It may require legislation, delegated rules, lawful regulatory or judicial action, or refusal to procure, depending on the field. Pakistan has not thereby adopted the EU Act's prohibited-practices list; Articles 5–6 offer comparative choices, not a domestic schedule (European Parliament & Council of the European Union, 2024).

These are oversight responses, not labels of virtue. A beneficial system may still require prior assessment, and high consequence does not mean certain harm. Classification determines who must prove what, before which body, with which records and controls. It must record residual risk: a tier should not fall merely because untested or unfunded safeguards are listed.

Lifecycle modifiers and institutional reality

Chapters 2 and 3 show why technical complexity is a poor proxy for consequence. A simple rules engine can exclude a household at scale when its source registry is stale, fragmented or difficult to correct. A highly complex model can remain low consequence in an isolated research environment. The protocol must therefore examine the complete chain: source records; collection and linkage; labels and training; model and threshold; procurement and integration; user presentation and reliance; final action; monitoring, incident response and downstream reuse.

Four groups of modifiers are particularly important. **Data risk** includes uncertain provenance, missingness, duplication, outdated identity or household records, proxy variables, unrepresentative languages or regions, and absence of a workable source-record correction process. Model accuracy cannot repair a false input, and correction of an output alone cannot prevent the source error recurring. **Model and measurement risk** includes labels that reproduce past discretionary practice, inappropriate comparators, weak calibration, unequal error, out-of-distribution use, uncertain generative output and a threshold chosen without reference to the relative costs of false positives and false negatives.

Procurement and dependency risk arises where the institution cannot inspect documentation, reproduce a result, control an update, obtain incident support, maintain service continuity or retrieve and delete its data at exit. Supplier assurances and intellectual-property claims may be managed through protected access, but an institution should not lower a tier on evidence it cannot test. NIST notes that third-party data, software and hardware can produce misaligned risk measures and opacity between developer and deployer (Tabassi, 2023). The deploying institution's loss of practical control is itself a reason for stronger conditions, not a transfer of public responsibility to the vendor.

Security and cumulative risk includes poisoning, evasion, model extraction, compromised dependencies, unauthorised access, insecure interfaces and manipulation of outputs. Risk increases where a common model or database becomes critical infrastructure or a single error travels from identity to benefits, credit, policing or employment. Frequency matters as well as single-event severity: small denials repeated against the same language group, district or documentation status can produce systemic exclusion. Institutional capacity is therefore a modifier. Missing logs, inadequate specialists, no independent testing budget, an inaccessible complaint channel or no manual fallback may turn a theoretically reversible error into an irreversible one.

Uncertainty must not be coded as safety. Where the proponent cannot identify training or evaluation data, affected population, subgroup performance, security assumptions, human reliance or a viable review route, the system should remain in the higher plausible tier, be conditionally tested, or not proceed. The evidential burden should rest principally on the institution seeking to expose people to the use, because affected persons cannot normally discover vendor-controlled facts before deployment. Classification is repeated when purpose, data, model, threshold, population, integration, supplier, security environment, human role or observed outcome materially changes. It is not a one-time procurement entry.

Research, security, emergencies and controlled experimentation

Neither innovation nor institutional urgency justifies a categorical exemption. **Research** conducted offline with synthetic or properly protected data and no effect on persons may remain under ordinary research, ethics, data and security governance. The tier rises when researchers use identifiable or sensitive records, recruit vulnerable participants, test in live services, return individual predictions, or permit an output to affect treatment or entitlement. Ethics approval does not itself confer administrative power, and publication value does not displace consent, confidentiality, equality, safety or source-specific legal authority. UNESCO's Recommendation supports proportionate ethical impact assessment throughout the lifecycle and a human-rights-centred approach, but it is an

international recommendation rather than directly enforceable Pakistani AI legislation (UNESCO, 2021).

Policing, defence and national security require context-specific legal treatment; a security label is neither an automatic prohibition nor an immunity from classification. Legitimate secrecy may justify a non-public register or controlled disclosure, but it strengthens the need for internal records, named authorisation, necessity and proportionality analysis, access controls, adversarial testing, audit, evidential preservation and review by a lawfully competent body. Where an output can initiate surveillance, targeting, detention or evidentiary use, its consequence cannot be reduced merely because a human receives the alert. Restrictions may specify that a match or prediction is an investigative lead requiring corroboration, not proof of identity, guilt or future conduct.

An **emergency** may change the urgency and available legal powers, but not the need to identify them. Emergency use should state the triggering event, statutory or constitutional basis, necessity, geographic and temporal scope, responsible official, permissible data and decisions, human override, incident route, continuity arrangements and sunset date. Contemporaneous logging and prompt retrospective review are essential where prior consultation is impossible. Data or models assembled for a flood, epidemic or cyber incident should not quietly migrate into routine policing, eligibility or employment after the emergency ends.

Sandboxes, pilots and controlled trials are governance devices, not law-free spaces. Pakistan has sectoral experience with limited live testing through the Securities and Exchange Commission of Pakistan's regulatory sandbox, while the *National Artificial Intelligence Policy 2025* proposes AI sandboxes (Securities and Exchange Commission of Pakistan [SECP], n.d.; MoITT, 2025). Neither creates a general exemption from constitutional, sectoral, consumer, data, procurement or criminal law. A trial needs an approved protocol, bounded purpose and sample, time limit, suitable notice, safeguards where consent is impracticable, complaint and compensation routes, preservation and incident duties, stop rules, independent evaluation, and rollback or data-disposal plans. Hidden expansion must be prohibited. Operational transition requires current evidence, reclassification and consequence-appropriate authorisation.

A consequence-classification protocol

The following minimum record converts principle into an auditable decision. It should be completed before procurement or internal development, signed by the responsible institution rather than the supplier, and retained as a living record.

Classification should also be performed at the level of the **deployment configuration**, not merely the purchased product. The same foundation model, matching engine or anomaly detector may sit in several workflows with different data,

thresholds, interfaces and consequences. A department that licences a general platform should therefore create a separate entry for each material use and for each institutional integration that changes who is exposed or what follows from an output. Common infrastructure may justify shared technical evidence, but it does not eliminate the need for a sector authority to establish its own lawful purpose, consequence, human role and remedy. Where one component serves several institutions, the protocol should name a lead custodian for shared risks and a responsible decision-maker for each downstream act. This prevents a platform-wide approval from becoming an unnoticed delegation of sectoral power, while allowing common validation, security testing and procurement assurance to be reused when their scope and continuing relevance are documented.

Table 5.1 | Classification protocol

Protocol field	Required record and classification question	Effect on oversight
System identity, status and function	Owner, version, vendor; announced, proposed, procured, piloted, operational, suspended, discontinued or unverified status; prediction, classification, ranking, recommendation, matching, generation, anomaly detection or action; point of influence in the workflow	Prevents product labels and documentary uncertainty from concealing actual use
Authority, purpose and sector	For public action, enabling provision, competent body and official; for private or regulated use, lawful basis and any required licence or permission; lawful objective; intended and prohibited uses; federal, provincial, local and sectoral setting	No adequate authority or lawful basis means no consequential pilot or operational deployment; sector law determines who may impose conditions
Affected population and interests	Persons directly and indirectly exposed; children or groups facing documentation, language, disability, poverty or access barriers; rights, safety, services, employment, finance or public power affected	Vulnerability or serious protected interests creates a presumptive higher floor
Consequence, scale and reversibility	Best and worst credible outcomes; volume, geography, frequency and duration; error persistence, correction time, downstream propagation and manual fallback	Severe, large, cumulative or hard-to-reverse effects elevate the tier
Data and inference	Sources, provenance, sensitivity, linkage, retention, labels, proxies, population coverage, quality and correction route	Sensitive, weak or uncorrectable records require stronger control or restriction
Human role and review route	Information shown, competence, time, authority and technical capacity to depart; notice, correction, reconsideration, appeal and urgent relief	Nominal review is treated as de facto automation and an absence of meaningful human control; genuine review may mitigate only relevant decisional risk

Protocol field	Required record and classification question	Effect on oversight
Model, validation and security threat	Intended conditions, metrics and error distribution; uncertainty; drift; attack surface, adversarial threats, access controls, logs, testing and incident route	Missing or non-transferable evidence prevents a lower tier; high-consequence exposure requires prior security assurance
Vendor and institutional dependence	Components, subcontractors, update control, audit and evidence access, continuity, portability, data return/deletion, skills and budget	Lock-in, opacity or absent institutional capacity elevates conditions and requires an exit plan
Proposed tier and decision	Ordinary, enhanced, prior assessment, conditional, restricted or prohibited; safeguards, residual risk, named approver, reasons, review date and material-change triggers	Makes the duty, responsible actor and basis for reclassification reviewable

The protocol's decisive discipline is attribution. It does not declare "AI risk" in the abstract; it links a defined use to authority, affected interests, evidence, safeguards, an accountable approver and a route of review. Courts, auditors, ombuds institutions and sector regulators remain free to apply their own lawful standards, and a low administrative tier cannot oust their jurisdiction. Used in that manner, consequence classification becomes the entry point to the federal-sectoral model developed in Section 5.2: common questions, differentiated controls, and responsibility retained by the institution legally empowered to act.

5.2 Allocation of Institutional Responsibility in a Federal-Sectoral Model

Institutional fit rather than organisational fashion

Pakistan does not face a binary choice between one artificial-intelligence regulator and institutional self-government. AI is not a freestanding constitutional subject. Its regulation follows the activity in which it is used: federal banking and telecommunications, provincial health and school administration, criminal law and evidence, municipal services, procurement, employment, professional practice, or a government's internal administration. The Constitution gives Parliament exclusive authority over the Federal Legislative List; preserves concurrent competence over criminal law, criminal procedure and evidence; and otherwise assigns unenumerated matters to the provinces. It also requires each province to devolve political, administrative and financial responsibility to elected local government (*Constitution of the Islamic Republic of Pakistan*, 1973, arts. 140A, 142). The first design rule is therefore *competence before convenience*: a coordinating body cannot acquire jurisdiction merely because a system is digital, technically complex or supplied across Pakistan.

That rule does not preclude common institutions. The Digital Nation Pakistan Act 2025 establishes a National Digital Commission (NDC), chaired by the Prime Minister and including the four Chief Ministers, relevant federal ministers and heads of major federal bodies, and a Pakistan Digital Authority (PDA). The NDC approves the National Digital Masterplan and coordinates federal, provincial and sectoral bodies. The PDA develops and monitors that Masterplan, coordinates across governmental levels, reviews public-sector projects with digital components, and may set standards in areas assigned by the NDC. Section 8(g) requires it to avoid overlap with another regulator's mandate and conflict with existing law; section 8(h) preserves each designated data custodian's control and responsibility (*Digital Nation Pakistan Act*, 2025, ss. 3–5, 8, 11–12). The statute supplies a coordination spine, not a transfer of every sector's decisional authority. Chief Ministers' NDC membership provides an intergovernmental forum; it does not displace constitutional competence.

The approved National AI Policy 2025 reinforces, but does not erase, that distinction. It envisages an AI Council, chaired by the federal Minister for IT and Telecommunications, and a Policy Implementation Cell in MoITT (MoITT, 2025, "Policy Implementation Mechanism"). They are policy-implementation machinery, not statutory cross-sector regulators. The Council's proposed membership includes provincial Chief Secretaries, the PDA Chairman and representatives of academia, industry, civil society and citizen advocacy groups. This offers a forum for coordination, but participation does not transfer the participants' legal powers to the Council. An Act, subordinate legislation, statutory direction, approved executive policy, draft proposal, intergovernmental agreement and model standard do not have identical force. Policy cannot enlarge legislative competence, confer sanctions on an advisory council, or displace statutory regulators, provincial executives, ombuds institutions or courts.

Comparing four oversight models

A **centralised model** places cross-sector standard-setting, licensing, investigation and enforcement in one specialist authority. Its attraction in Pakistan is practical. Scarce technical staff can be pooled; one register can reveal duplicated acquisitions and vendors; shared testing can serve institutions unable to maintain it; and national reporting can expose an error travelling between identity, benefits, finance and policing. Central purchasing leverage may secure audit access unavailable to a small department. Central direction has its strongest case for federal digital public infrastructure, interoperability, national cyber coordination and genuinely cross-jurisdictional systems.

Yet full centralisation would carry four costs. First, it may confuse nationwide technological reach with federal legal competence. A federal classifier used by a provincial hospital remains embedded in healthcare delivery and professional duties that cannot be reassigned by nomenclature. Secondly, a general AI office may understand

model testing while lacking the clinical, educational, financial or policing knowledge needed to decide acceptable error and remedy. Thirdly, concentrating registers, expertise and approval in one office can create a bottleneck and a single point of administrative, political and cybersecurity failure. Fourthly, it may duplicate existing regulators and permit responsibility shifting: a department could defend an unlawful decision by saying that "the AI authority approved the model". Research based on sixteen expert interviews identifies fragmented institutions, resource constraints and dependence on external technologies in Pakistan's cyber governance (Ahmed et al., 2025). Adding an overlapping institution without settled decision rights could reproduce those conditions rather than cure them.

A **sectoral model** leaves supervision with existing ministries, provincial departments, regulators and professional bodies. Its strength is proximity to the regulated consequence. The State Bank of Pakistan can assess credit, payment and banking-risk uses within banking law; the Securities and Exchange Commission of Pakistan can address systems used in securities, insurance and non-bank finance; the Pakistan Telecommunication Authority can apply telecommunications law and subscriber protections; and drug, medical and healthcare bodies can distinguish administrative optimisation from clinical decision support (*State Bank of Pakistan Act*, 1956; *Pakistan Telecommunication (Re-organization) Act*, 1996; *Securities and Exchange Commission of Pakistan Act*, 1997; *Drug Regulatory Authority of Pakistan Act*, 2012). The Federal Board of Revenue understands the legal difference between selecting a return for audit and raising a tax demand. Provincial home departments and police organisations control operational policing subject to criminal-process law, while provincial health, education, land and local-government institutions possess the service records and frontline staff through which consequences occur. Sector supervision therefore connects validation thresholds, professional judgement and remedies to the legal purpose of the service.

Pure sectoralism nevertheless fragments the picture. Institutions may define "high consequence" inconsistently, vendors may exploit divergent documentation requirements, and a regulator may miss cumulative exclusion produced by several lawful-looking decisions. Capacity is unequal between sectors and provinces. A deployer may regulate itself where its department lacks an independent technical team or external regulator. Foundation models, identity services and data brokers also cross silos. Sectoral autonomy therefore needs common minimum methods, referral duties and shared technical support.

A **coordinated model** preserves sector mandates while a national body supplies taxonomy, templates, incident channels, information exchange and consolidated reporting. It is less constitutionally disruptive and allows standards to be adapted to local institutions. Its weakness is the familiar "everyone coordinates, no one decides" problem. A forum without named owners, deadlines, records and escalation routes can produce

meetings rather than control. Conversely, a coordinator that issues nominally "technical" requirements affecting eligibility, policing or professional practice may exercise coercive power without the legislation and procedural safeguards attached to a regulator.

The most suitable **hybrid model** is consequently neither a compromise for its own sake nor a copy of one foreign architecture. It separates functions that require consistency from decisions that require lawful domain authority. The national layer maintains common infrastructure, vocabulary and learning; the competent sectoral or governmental body authorises and supervises the use; the deploying institution remains answerable for each decision; data custodians remain responsible for source records; CERT structures address cybersecurity; procurement and audit bodies test expenditure and contract performance; and ombuds, rights bodies and courts retain independent review. Different uses may still justify different emphases. A federal payments infrastructure may sit closer to the national end of the spectrum, while an attendance-risk tool in one provincial school system should remain primarily provincial and educational. "Hybrid" is a method of allocation, not a licence to give every body concurrent jurisdiction.

A bounded national coordinating function

The national coordinating layer should undertake seven tasks. It should maintain a technology-neutral consequence taxonomy and interoperable register schema; issue model templates for impact assessment, incident reporting, audit access and decommissioning; operate a secure information-exchange mechanism; aggregate performance and complaint data without unnecessary central pooling of personal records; provide shared validation, procurement and training support; sponsor independent research and cross-provincial learning; and report publicly to Parliament, provincial assemblies and the NDC on implementation and unresolved gaps. A shared register may be physically federated: the PDA can maintain searchable metadata while the responsible authority preserves detailed assessments, security records and case files. This design would preserve the responsibility of designated data custodians under section 8(h) of the Digital Nation Pakistan Act 2025 while avoiding unnecessary replication of operational databases (Digital Nation Pakistan Act, 2025, s. 8(h)).

The PDA is the logical secretariat for these common functions where they fall within the Digital Nation Pakistan Act, any duly approved National Digital Masterplan or another valid supporting instrument. The National AI Policy's AI Council can supply specialist and participatory advice, but its terms of reference should state its relationship to the statutory NDC and PDA, and to the National Data Governance Council proposed in the draft policy. It should not become a parallel licensing authority by policy assertion. Binding requirements must identify their source: a PDA regulation or standard within the 2025 Act; a federal or provincial procurement rule; a sector

regulator's enabling statute; a provincial law or direction; a consensual intergovernmental instrument; or primary legislation. The same discipline applies to the Policy's proposed AI Directorate, which its action matrix associates with a contemplated personal-data commission. Unless and until legislation creates that commission and assigns functions to it, the proposal cannot be treated as an operating statutory regulator (MoITT, 2025, p. 23).

Several powers should remain outside the coordinator's ordinary role. It should not decide whether an individual receives a benefit, is admitted, investigated, licensed or disciplined; replace a clinician, tax officer or other legally responsible professional; determine civil or criminal liability; decide complaints against itself or a sector body; or compel provinces on residual subjects without a constitutional or statutory route. Nor should it centralise source data as a condition of oversight. It may flag non-conformance, refer a case to the competent regulator, recommend suspension and publish aggregate findings. Coercive investigation, penalty, prohibition or compulsory disclosure requires an express legal basis and the associated opportunity to be heard.

Sectoral, provincial, local and independent mandates

The sector supervisor should translate common controls into its field's risk language. It defines the lawful objective, validation population, prohibited proxies, professional involvement, operating threshold, notice, complaint route and suspension conditions. The service-owning institution remains the accountable deployer: it selects data and vendors, trains staff, logs operation, gives reasons, corrects records and enables a human reviewer to depart from the output. Vendor certification may inform but cannot replace that judgement. Without a specialist regulator, a department should use an authorisation committee separated from the project team and, for high-consequence uses, external technical, rights and audit scrutiny.

Cross-cutting bodies retain their own mandates. Federal PPRA and provincial procurement authorities govern their respective procurements; a PDA model clause cannot itself waive procurement law, and a procuring agency remains responsible for an intelligible specification and lawful evaluation (*Public Procurement Regulatory Authority Ordinance*, 2002; *Public Procurement Rules*, 2004). The Auditor-General's constitutional audit extends to the accounts of the Federation, provinces and bodies established by either, while legislative Public Accounts Committees supply political follow-up (*Constitution of the Islamic Republic of Pakistan*, 1973, arts. 168–171). Audit should examine value for money, control design and contract delivery, but should not purport to decide the legality of an individual detention or clinical conclusion.

Cybersecurity follows another specialised network. The published Computer Emergency Response Team Rules 2023 describe provincial and other sectoral CERT constituencies, with National CERT coordinating threats not successfully resolved by

the relevant sectoral CERT and widespread attacks (*Computer Emergency Response Team Rules*, 2023, rr. 5, 9–10, 15–16). Rule 1(3), however, makes commencement dependent on a further Gazette notification, and none was identified in the official sources reviewed by the cut-off. An AI incident involving poisoning, model theft or compromise should therefore use that network where and to the extent lawfully operative, while the service owner's contract and incident plan must allocate action despite commencement uncertainty. Technical routing never substitutes for notice to a sector regulator or affected person when legal interests are harmed.

Review institutions must be connected without being subordinated. The Wafaqi Mohtasib addresses maladministration by federal agencies, with provincial and specialised ombuds institutions acting under their own instruments (*Establishment of the Office of Wafaqi Mohtasib (Ombudsman) Order*, 1983; *Federal Ombudsmen Institutional Reforms Act*, 2013). The National Commission for Human Rights may inquire into human-rights violations or negligence in their prevention, seek evidence and recommend action (*National Commission for Human Rights Act*, 2012, ss. 9, 12–18). Professional bodies address practitioner standards; tribunals exercise jurisdiction conferred by sector law. High Courts retain Article 199 jurisdiction over unlawful public action and rights enforcement, while the Federal Constitutional Court exercises the jurisdictions introduced by the Twenty-Seventh Amendment, including relevant appeals and intergovernmental disputes (*Constitution of the Islamic Republic of Pakistan*, 1973, arts. 175E–175F, 199). Section 29 of the Digital Nation Pakistan Act cannot simply be treated as ousting constitutionally vested review; its reach is for the courts to determine (*Digital Nation Pakistan Act*, 2025, s. 29). A coordinator may preserve evidence and publish referral maps, but cannot screen those routes.

Local institutions require more than nominal consultation. Municipal offices, district health authorities, schools, land-record centres and local complaint desks often create the data, communicate the result and detect error first. Where an AI system changes these functions, its assessment must record which local body owns each source field, who can correct it, what offline channel exists and who continues service during failure. National or provincial dashboards that omit local workload and complaint data will systematically overstate success. Article 140A also makes it inappropriate to bypass a legally devolved local function merely because a provincial or federal vendor offers a common platform.

Responsibility should therefore be allocated as a set of **non-transferable bundles**, rather than a long list of organisations invited to a committee. The institution holding a coercive or distributive power must own the legality of its use, the decision criteria, the final reasons and the remedy. The data custodian must own the accuracy, provenance and correction of its authoritative record. The system owner must maintain configuration, access, logging, security and continuity; the procuring agency must secure evidence, change-control and exit rights; and the competent supervisor must possess

information and enforcement powers sufficient for review. One organisation may hold several bundles, but no bundle should disappear between contract schedules or memoranda of understanding. Where responsibility and control diverge for example, a province bears service liability but a federal platform alone can change the matching rule the intergovernmental instrument must confer the necessary access and change route, allocate incident decisions and fund compliance. If lawful control cannot be aligned with responsibility, the integration should not proceed. This discipline also protects reviewers: an auditor, ombuds institution or rights commission should not be described as sharing operational ownership merely because it later inspects the system.

Formal intergovernmental procedure

Pakistan can formalise coordination by using, and clarifying, the institutions already created rather than immediately adding another authority. The AI Council envisaged by the 2025 Policy should be notified with written, publicly available terms of reference as the specialist advisory forum serving the NDC and PDA architecture. Each province should have a named voting representative and technical alternate; relevant sector regulators should join agenda-specific working groups; and local government, affected-community, disability, linguistic, professional, academic and civil-society representatives should participate where the proposed standard affects them. UNESCO's Recommendation supports inclusive participation and oversight as elements of ethical AI governance (UNESCO, 2021). Rights commissions, auditors and ombuds institutions should ordinarily attend as independent observers or evidence providers rather than be made collectively responsible for an operational decision they may later review.

Procedure is as important as membership. A permanent secretariat should circulate an agenda, jurisdiction note, draft instrument and fiscal estimate in advance; maintain a conflict-of-interest register; record provincial and sectoral objections; publish non-security minutes and an action tracker; and classify every output as binding law, statutory instrument, policy direction, model rule, interoperability standard, funding condition or non-binding guidance. Consensus should be sought for model standards. Majority agreement in an administrative council cannot transform a recommendation into law for a dissenting province. A province adopting a common standard should identify the provincial Act, rule, cabinet decision, contract or departmental direction that gives it effect.

For a national system operating in a provincial sector, the process should require a joint jurisdiction and impact memorandum before procurement: the federal interest and legal basis; the provincial function and data custodian; local implementation duties; validation for the province's population and languages; complaint and judicial routes; funding for staff and correction; and an exit arrangement. Common minimum

outcomes may coexist with provincial implementation schedules and stricter safeguards. Financial participation should carry disclosed milestones, not an implied surrender of constitutional competence. Where the Federation confers duties on a province under Article 146, the constitutional text itself anticipates agreement on additional administrative cost; where a province entrusts functions to the Federation under Article 147, provincial-assembly ratification is required within sixty days. Parliament may legislate on an otherwise provincial matter for consenting provinces under Article 144 (*Constitution of the Islamic Republic of Pakistan*, 1973, arts. 144, 146–147).

Disputes should move through a recorded ladder: technical reconciliation by the relevant working group; a jurisdiction opinion from the federal and provincial law departments; principals' negotiation in the AI Council or NDC; and referral to the Council of Common Interests only where the matter falls within its constitutional remit over Part II of the Federal Legislative List. That remit includes national planning and economic coordination, federally established regulators, legal and medical professions, and higher-education standards, but it is not a general power over every digital service (*Constitution of the Islamic Republic of Pakistan*, 1973, arts. 153–154 and Fourth Schedule, pt II). Unresolved disputes between governments ultimately remain within the Constitution's judicial arrangements. This ladder makes disagreement visible and prevents a technical secretariat from deciding a question of federal competence by default.

Federal-sectoral lifecycle responsibility matrix

The following matrix assigns a lead rather than merely listing interested bodies. "Deploying authority" means the institution legally responsible for the service or regulated activity; it remains accountable even where a vendor operates the system.

Table 5.2 | Lifecycle responsibility matrix

Lifecycle function	Accountable lead	National/common role	Sectoral, provincial and local role	Independent check and required record
Lawful purpose and authority	Deploying authority's head, advised by its law office	PDA supplies a jurisdiction template; NDC facilitates cross-government issues	Competent department/regulator confirms sector power; province approves provincial function; local body identifies devolved duty	Courts retain legality review; signed authority-and-purpose memorandum

Lifecycle function	Accountable lead	National/common role	Sectoral, provincial and local role	Independent check and required record
Consequence classification and register	Deploying authority proposes; competent regulator or departmental authoriser confirms	PDA maintains common taxonomy and federated national metadata register	Provincial/sector register holds detailed file and local service map	Public status, tier, owner, legal basis and review route; justified security redactions
Data and integrated impact assessment	Deploying authority and lawful data custodian	PDA provides common data/AI assessment method and cross-system dependency analysis	Sector experts test relevance; province and local source owner verify population, language, accessibility and correction	NCHR/rights or equality expertise for high-consequence uses; approved assessment and dissent log
Procurement and funding approval	Procuring agency; relevant finance/planning authority for funds	PDA issues model AI clauses and identifies reusable capability	Federal or provincial PPRA regime governs process; sector body writes outcome and validation specifications	Procurement review, AGP/internal audit; business case, lifecycle cost and contract-access schedule
Validation and cybersecurity testing	Competent sector authoriser accepts evidence; deployer owns residual risk	Shared test facilities and benchmarks; National CERT supports national or escalated threats within an applicable mandate and operative route	Independent tester and data custodian test local data, workflow and security; sector/provincial CERT participates where constituted and lawfully operative	Validation report by group and location; threat model, limitations and remediation record
Pilot or operational authorisation	Statutorily competent regulator, minister, department or provincial authority	PDA authorises only where the 2025 Act/Masterplan validly assigns that function; coordinator records decision	Sector/province sets conditions, thresholds, professional supervision and service-continuity plan	Written, time-limited authorisation; responsible officer, complaint route, stop and exit criteria
Operation and human decision	Deploying authority and named official/professional	Common logging, notice and training standards	Frontline/local body gives notice, records intervention, corrects source record and maintains non-digital access	Case-level reasons, input provenance, human action and final decision log

Lifecycle function	Accountable lead	National/common role	Sectoral, provincial and local role	Independent check and required record
Cyber/AI incident response	System owner contains harm and preserves evidence	National CERT coordinates an escalated or widespread cyber incident where its mandate and operative legal route apply; PDA correlates cross-system AI incidents	Sector/provincial CERT acts where constituted and lawfully operative; regulator and data custodian address sector and record consequences	Incident file, regulat or/affected-person notice where legally required, continuity and recovery record
Monitoring and material-change review	Deploying authority; sector regulator supervises	PDA aggregates comparable performance, vendor and incident indicators	Province/local service reports errors, reliance, complaints and distributional effects	Periodic monitoring report; reauthorisation after material change
Complaint and individual redress	Deploying authority provides correction and genuine reconsideration	National portal may route, track and publish aggregates, but not adjudicate	Regulator/department hears appeal; local and offline intake remains available	Ombuds, tribunal, commission, professional body or court as law provides; complete review record
Audit, rights and legality review	Institution depends on mandate: AGP/audit, NCHR, ombuds, tribunal or court	Coordinator supplies preserved records and system-level analysis	Provincial audit/assembly and rights institutions examine their jurisdictions	Findings remain institutionally independent; disclosure schedule and compliance response
Evaluation, renewal and decommissioning	Service-owning government or regulator decides; finance body controls continued funding	PDA reports comparative results and cross-government lessons	Province/sector assesses lawful outcome and alternatives; local body verifies service and record transition	Legislature/auditor reviews expenditure; renewal decision, data return/deletion, vendor exit and remedy completion

The matrix prevents three recurring evasions. National approval does not legalise an action outside the deployer's authority; sector approval does not discharge the PDA's valid common-standard functions; and outsourced operation does not make the vendor the public decision-maker. For every system, one institution must own the decision, one custodian must own each authoritative record, one body must hold the authorisation power, and every independent review route must remain identifiable. Coordination succeeds when it makes those allocations visible and supplies capabilities that institutions cannot efficiently build alone. It fails when shared responsibility becomes untraceable responsibility.

5.3 Ex Ante Controls: Registers, Impact Assessment, Authorisation, and Procurement Gates

Ex ante governance must begin before institutional enthusiasm, sunk procurement costs and technical integration make refusal politically or financially difficult. It is therefore not a single approval given to "the AI system". It is a sequence of decisions about a defined use: whether a public body has authority to pursue the purpose; whether automation is necessary and proportionate to that purpose; whether the relevant data and model are fit for the affected population; whether the institution can supervise and remedy the use; whether the procurement preserves public control; and, finally, whether the configured system may be placed into production. A department may lawfully purchase analytical software but lack authority to use its output to determine eligibility, discipline an employee or flag a person for investigation. Conversely, a statutory power to make the underlying decision does not by itself authorise new data linkage, undisclosed profiling or delegation of official judgement to a vendor. Purchase approval and deployment authorisation must accordingly remain distinct.

Pakistan already possesses parts of this architecture. The approved National Artificial Intelligence Policy 2025 an executive policy, not a statute calls for public registration of public-sector AI, lifecycle evaluation and impact assessment of high-impact systems, human oversight, third-party audit and regulatory sandboxes (Ministry of Information Technology and Telecommunication [MoITT], 2025). Sections 8, 11 and 12 of the Digital Nation Pakistan Act 2025 give the Pakistan Digital Authority (PDA) functions concerning standards, data governance, review of public projects containing digital components and implementation of the National Digital Masterplan, subject to the Act's safeguards against overlap with established regulators. The federal Public Procurement Rules 2004 already require procurement planning, competent-authority approval, generic specifications, announced evaluation criteria, records and post-award transparency (*Public Procurement Rules*, 2004). These are genuine legal footholds, but they are not a complete AI authorisation statute.

The distinction between proposals and enforceable duties also applies to data protection. The Personal Data Protection Bill 2023 is a draft text, rather than an enacted source of generally applicable rights or penalties (MoITT, 2023). Its proposals should be analysed separately from duties arising under the Constitution, sector legislation, cyber law, access-to-information law and administrative law. The authorisation model developed here draws on those existing sources and on the National AI Policy's assessment, oversight and audit commitments (MoITT, 2025, s. 3.2). Each binding requirement still needs a competent authority and a lawful instrument. A policy commitment to stronger safeguards cannot by itself create jurisdiction, sanctions or a new cause of action.

Registers: visibility before commitment

Every public body, and every regulated entity deploying a high-consequence use, should maintain a versioned internal register from the concept stage. A public entry should follow before procurement or pilot unless a field-specific exemption is justified. Registration is not certification and must not be presented as government endorsement. Its functions are limited but important: to prevent systems from remaining unknown to senior officials and regulators; identify who is answerable; connect procurement, assessment and incident files; and allow an affected person to find the institution and review route. Each use should receive a persistent identifier that survives changes of vendor or model. The register should state whether the use is announced, proposed, procured, piloted, operational, suspended, discontinued or unverified. A procurement notice is not proof of operational deployment.

The public entry should identify the responsible institution and senior owner; legal authority and lawful purpose; sector and decision function; vendor or in-house developer; affected population and geographical coverage; categories and sources of input data; the role of the output and of the human decision-maker; consequence tier; current status and relevant dates; validation and impact-assessment summaries; procurement route; complaint, correction and human-review channels; approving and reviewing bodies; and links to policies, audit summaries and material-change notices. Where the use affects individuals, the description should say in ordinary language whether the system classifies, ranks, predicts, matches, recommends, generates or triggers action. Merely publishing a product name, the label "AI-enabled", or an impenetrable model card does not meet the transparency purpose identified in the literature supplied for this chapter (Cotino Hueso, 2023).

The internal entry should add non-public detail: exact datasets and fields, provenance and transformation records, model and dataset versions, thresholds, integrations, architecture, threat model, known vulnerabilities, subgroup test results, complete contracts and subprocessor chains, data locations, retention rules, named operational staff, risk acceptance, contingency arrangements and the location of logs and preserved evidence. This two-layer design reconciles transparency with legitimate security and confidentiality. Withholding should occur at field level, not by making the whole system invisible. The record should identify the legal ground, anticipated harm, decision-maker and review date for every redaction; the unredacted dossier must remain accessible to the competent regulator, auditor and court under controlled conditions.

For federal bodies, Article 19A and the Right of Access to Information Act 2017 support a presumption of proactive disclosure concerning powers, policies, contracts and final decisions, while permitting legally justified exceptions and severance. The approved AI Policy supplies an immediate policy basis for a common register, and departments can establish internal inventories through administrative directions. PDA

can develop a common schema within a duly approved Masterplan or standards programme; sector regulators may require reporting where their enabling statutes permit. A legally mandatory, cross-sector public register covering private as well as public actors, backed by investigation and sanction, would require clear delegated power or primary legislation. Federal policy also cannot, by assertion, bind provincial departments in devolved sectors; compatible provincial registers should be joined through an agreed common schema rather than treated as branch offices of a federal database.

One integrated assessment, not a collection of forms

A composite assessment bringing together fundamental-rights, privacy and AI risk-management analysis is sound, provided integration does not dilute any component (Parlov et al., 2025). These comparative frameworks are design prompts, not Pakistani law. Pakistan should use one integrated AI and institutional impact assessment with proportionate annexes, rather than separate checklists whose conclusions never meet. The dossier should be initiated by the service owner, challenged by legal, data, security, procurement and domain specialists, and approved by the official who bears authority for the proposed use. High-consequence assessments should receive independent technical scrutiny; supplier assurance can inform, but cannot constitute, the public body's assessment.

First, the assessment must establish authority, purpose and necessity. It should identify the precise statutory or administrative power exercised, the institution and level of government possessing it, any delegation, the permissible data-sharing basis, and the reasons why the proposed use advances a lawful objective. It should compare the system with realistic alternatives: improved source records, additional staff, a simpler rule, sampling, professional decision support, or no intervention. The relevant counterfactual is not whether the model is accurate in isolation, but whether it improves lawful outcomes over current practice after false positives, false negatives, delay, exclusion, staff time and correction costs are counted. A claimed efficiency gain cannot cure absence of power.

Secondly, the assessment should examine data and model integrity together. For each material input it should record source, custodian, collection authority, original purpose, lineage, completeness, update cycle, regional coverage and known error. It should distinguish an incorrect model inference from a correct computation based on a wrong identity, land, tax, health or beneficiary record. It should test whether labels encode earlier discretionary or discriminatory practice and whether proxy variables reproduce geography, poverty, disability, gender, language, migration or documentation status. Validation must use the population, institutions, languages and operating conditions in which the system will be used, report uncertainty and confidence intervals where appropriate, and compare errors across materially affected groups. Aggregate "accuracy"

is inadequate where one error deprives a person of income, treatment, liberty or an opportunity. Threshold choice, calibration, override rate and the expected consequences of error are policy decisions requiring recorded reasons, not settings to be left to a vendor.

Thirdly, rights and accessibility analysis should trace the decision pathway. It should identify effects on life and liberty, dignity and privacy, equality, livelihood, education, health, property, expression, reputation and procedural fairness; specify notice, intelligible reasons, record correction and genuine human reconsideration; and test the route with people facing low literacy, disability, weak connectivity, distance, language barriers or urgent need. Consultation must be targeted to groups likely to bear error, not limited to technically sophisticated stakeholders. Where children, patients, detainees, employees, benefit recipients or other dependent populations are involved, apparent consent may not answer coercion or power imbalance.

Fourthly, the dossier should contain a system-specific cyber and resilience assessment: attack surface, poisoning and evasion, access controls, encryption, model extraction, supply-chain and API dependence, logging, backup, incident routes and safe failure. The CERT Rules 2023 are a published delegated instrument; rule 1(3), however, conditions commencement on a further Gazette notification, so their application requires confirmation of the operative legal position. The National Cyber Security Policy 2021 supports supplier assurance, security assessments and an information-security assurance framework, while remaining an executive policy (MoITT, 2021, ss. 3.5–3.6). The NCSC's Guidelines for Secure AI System Development add comparative guidance on security throughout design, development, deployment and operation (*National Cyber Security Centre, 2023; Computer Emergency Response Team Rules*, 2023). These controls do not prove model validity.

Finally, the assessment must test institutional and financial fit. It should name the professionals legally entitled to act, assess their competence and workload, show that they can inspect relevant records and depart from the output, and budget for monitoring, complaints, legal assistance, independent testing, incident response and exit. It should identify provincial and local roles where they create the records or deliver the service. It must also map vendor concentration, subcontractors, foreign hosting, licensing constraints, portability and the cost of maintaining a non-automated fallback. Proposed remedies correction, reconsideration, restoration, suspension and compensation should be designed before harm, not discovered after a complaint. A public summary should disclose the purpose, evidence, principal risks, mitigations, residual risks and approval, while confidential annexes preserve legitimately protected material. Documentation here is operational infrastructure for later accountability, not an archival exercise (Königstorfer & Thalmann, 2022).

Authorisation as staged proof

The following protocol separates gates that are often collapsed into one procurement approval. Failure at an early gate stops later expenditure; conditional passage must state measurable conditions, an owner and an expiry date.

Table 5.3 | Ex ante procurement and authorisation gates

Gate	Minimum evidence required	Accountable decision	Principal legal vehicle
0. Concept and register	Defined function, status, population, consequence tier, service owner and preliminary authority	Department or regulated-entity owner	Immediate internal direction; common register standard by PDA/sector regulator; legislation for universal publication and sanctions
1. Lawful need and assessment	Legal power, purpose, necessity and alternatives; integrated rights, equality, data, federal-fit, accessibility, capacity and remedy assessment	Competent administrative authority with legal and domain sign-off	Existing administrative and sector powers; formal sector rules where approval affects regulated firms
2. Market and tender	Outcome-based specification, whole-life cost, evaluation method, vendor due diligence, mandatory contract schedule and exit design	Procurement authority under an explicit delegation	PPRA/provincial procurement rules; AI standard procurement document or regulation
3. Technical and security acceptance	Local and subgroup validation; provenance; threshold justification; adversarial and security tests; logging; trained human oversight; corrected critical defects	Deploying authority, independent validator and relevant sector/cyber body	Contract acceptance conditions; applicable security policy and contractual controls; operative CERT and sector requirements
4. Production authorisation	Named responsible official, final configuration and version, notice and review channel, continuity plan, register update and time-limited approval	Official legally responsible for the service; regulator where its statute requires	Existing decision and licensing powers; primary legislation where no authority exists to impose a binding prior approval
5. Scale or material expansion	Pilot results, incident and complaint record, distributional outcomes, updated assessment and proof that conditions were met	Fresh decision by the original approving bodies	New or amended authorisation; never automatic conversion from pilot or contract award

The signatories should not become a diffuse committee in which no one is responsible. The service owner certifies operational need; legal counsel states the source and limit of power; the data steward certifies provenance and correction arrangements; the security officer addresses applicable security policy and contractual controls and any lawfully operative CERT duties; the domain professional accepts the performance and error trade-off; the procurement officer preserves the terms; and the competent authority decides whether residual risk is acceptable. Independence is required where consequence or uncertainty is high: the developer should not validate its own claims, and an

implementing consultant should not audit the same controls. If records, representative validation, override capacity or a workable remedy are missing, the answer at the gate is "not ready", not an undocumented acceptance of risk.

Procurement as a governance instrument

AI governance can fail before code exists because the tender purchases a product rather than defines a public problem; procurement is itself part of the legal control architecture (Vázquez Matilla, 2025). Federal Rule 8 requires realistic procurement planning; Rule 11 requires clear authorisation and competent approval; and Rules 10, 23 and 29 permit generic performance specifications, standards, contract conditions and pre-announced evaluation criteria. The amended definition of the "most advantageous bid" and "value for money" allows quality, reliability, after-sales service, upgradeability and whole-life cost to matter, rather than reducing evaluation to the cheapest licence. Pre-qualification under Rules 15–17 can test technical, legal, managerial and financial capacity. Complex AI may justify a two-stage procedure under Rule 37, but dialogue must remain equal and evaluation criteria transparent. Comparable provisions should be implemented through each provincial procurement regime rather than assuming the federal rules apply.

An AI procurement schedule should make the assessed use, prohibited uses, performance measures and acceptance tests contractually precise. It should require disclosure of training and validation data provenance; model, dataset and software versions; known limitations; threshold and calibration evidence; subgroup and local-language results; security architecture; energy and infrastructure dependencies where material; and all model providers, cloud hosts and subprocessors. The supplier should warrant authority to license relevant components and disclose material dependence on third-party terms. Evaluation should award points for correction, interoperability, portability, inspectability and safe degradation, and should price monitoring, retraining, hosting, human review, audit, migration and decommissioning across the contract life. Promotional benchmarks on an undisclosed population are not acceptance evidence.

The contract must then preserve the evidence needed by the institution, regulator, auditor, ombuds body and court. It should require decision and access logs, change histories, incident and human-intervention records, reproducible test artefacts and retained copies of every production version. Audit rights should extend to relevant subcontractors and permit independent testing, working papers and secure access to otherwise confidential technical material. Intellectual property may justify a controlled inspection room, confidentiality undertaking or escrow; it cannot give the vendor a veto over procedural fairness. Rule 46's minimum five-year procurement record is a floor for procurement proceedings, not necessarily the correct retention period for decision

evidence. The period should also account for sectoral record law, limitation periods, pending complaints, audit and litigation holds.

Security terms should specify vulnerability disclosure, patch and notification times, evidence preservation, cooperation with nCERT and the sector regulator, penetration and adversarial testing, access revocation and responsibility for compromised dependencies. Change-control clauses should require advance notice and approval for material changes to model, data, threshold, purpose, hosting, subcontractor or interface; an online vendor update must not silently alter an authorised public decision. Service levels should measure legally relevant outcomes, not merely uptime.

Exit provisions are equally substantive. They should permit suspension or termination for failed acceptance, unlawful use, material bias, serious incident, loss of regulator access, unapproved change, insolvency or repeated service failure; maintain a safe manual or alternative service; require export of data, logs, configuration and documentation in usable formats; secure return or verified deletion of copies; preserve evidence subject to legal hold; and provide transition assistance, escrow or replacement rights where dependency is acute. Liability allocation, indemnity, insurance, service credits and dispute resolution can distribute loss between buyer and supplier, but cannot contract away the public body's statutory or constitutional responsibility. PPRA can issue an AI-specific standard procurement document or model schedule under its existing quality and procedure functions; agencies can include consistent clauses now under Rule 23. Primary legislation is not required for ordinary contractual rights, although legislation may be needed to compel non-party model providers or override confidentiality against regulators and courts across all sectors.

Pilots, sandboxes and conditional authorisation

A pilot is a bounded evidential exercise, not a smaller deployment and not a device for avoiding procurement, rights or professional law. Pakistan has usable domestic precedents: the Securities and Exchange Commission of Pakistan's (SECP) Regulatory Sandbox Guidelines 2019 provide controlled testing within its financial jurisdiction, and the State Bank of Pakistan (SBP) issued Regulatory Sandbox Guidelines in 2025. The National AI Policy envisages integrated AI sandboxes assisted by Centres of Excellence (SECP, 2019; SBP, 2025a; MoITT, 2025). These examples demonstrate institutional feasibility, not a general power to waive law. Each regulator may relax or clarify only requirements within its enabling authority. It cannot waive constitutional protections, another regulator's statute, a provincial competence, criminal procedure or a person's right to seek a remedy. Scholarship on sandboxes likewise cautions that experimental permission does not displace ordinary liability rules (Riccio, 2025).

The authorisation should pre-register the hypothesis, lawful basis, regulator and accountable sponsor; model and data version; geography, duration, maximum

participants and transactions; inclusion and exclusion criteria; baseline and evaluation measures; notice or consent arrangements; human fallback; complaint and correction route; security controls; independent evaluator; prohibited actions; incident and pause thresholds; and exit method. Recruitment should not select only convenient, connected users and then claim national validity. Testing must cover relevant languages, regions, disabilities and administrative conditions, while preventing vulnerable populations from carrying irreversible experimental risk. Where a pilot touches benefits, health, education, employment, policing or liberty, it should ordinarily operate in shadow or advisory mode until safety is established; the existing lawful process remains available and no adverse action should rest solely on an experimental output.

Conditional authorisation should state what may occur, not merely what remains to be studied. Conditions must be objective for example, an error ceiling, zero unresolved critical vulnerabilities, a maximum caseload, mandatory second review for adverse outcomes, weekly incident reporting and a fixed expiry. Crossing a harm or security threshold should pause the use automatically. Pilot data should not be repurposed for product training or another agency without a fresh lawful basis and assessment. Participants must receive notice of material incidents and retain ordinary complaint and compensation routes. Results should be reported even when unfavourable, including distributional errors, overrides, complaints, incidents, cost and operational burden.

Federal Procurement Rule 42 permits negotiated tendering for supplies manufactured purely for a particular research, experiment, study or development, but requires reasons and justification on the record. That provision changes the acquisition method; it does not convert experimental subjects into unprotected users or establish authority for the intended administrative act. Nor may an "emergency" or proof-of-concept label conceal operational use. Any continuation, population expansion, linkage to another database or transfer from advisory to determinative use must return to the assessment and production gates. The default at expiry is cessation and orderly deletion or preservation, not silent rollover.

The resulting reform is deliberately staged. Departments and regulators can begin internal registers, integrated assessments, tender schedules, acceptance tests and bounded-pilot protocols under existing administrative, procurement and sectoral powers. PPRA, provincial procurement authorities, PDA, nCERT and sector regulators can standardise these measures through instruments issued within their respective enabling laws and with required approvals. Intergovernmental agreement is needed for common schemas and shared systems touching devolved services. Primary legislation is reserved for the gaps that contract and policy cannot safely fill: universal coverage, enforceable individual rights, regulator access notwithstanding vendor control, cross-sector sanctions, binding prior authorisation where no sector power exists, and durable coordination across jurisdictions. That separation makes the gates

implementable now while preserving the constitutional proposition on which they depend: technology may assist public administration, but lawful authority, official responsibility and access to remedy cannot be procured from a vendor.

5.4 Ongoing Assurance, Incident Governance, Transparency, Participation, and Redress

5.4.1 From authorisation to continuing assurance

Authorisation is a point in time; assurance is a continuing claim supported by evidence. A valid model can become unreliable as populations, source records, institutional practice, threats or surrounding software change. Post-deployment governance must therefore ask whether the whole socio-technical arrangement still achieves its lawful purpose for the authorised population, with acceptable error, security, human control and remedy. NIST identifies stale data and data, model or concept drift as reasons for corrective maintenance (Tabassi, 2023, Appendix B). This lifecycle problem matters in healthcare, benefit administration, policing, credit, education and revenue.

The legal and policy materials considered here supply several components of continuous assurance. The National AI Policy 2025 calls for a public register of public-sector AI, lifecycle evaluation of high-impact uses, regular audit, third-party accountability, redress and public participation (Ministry of Information Technology and Telecommunication [MoITT], 2025, s. 3.2 and "Policy Implementation Mechanism"). These remain policy commitments requiring lawful implementation. The published CERT Rules 2023 are a delegated instrument, but rule 1(3) makes commencement dependent on a further Gazette notification. Continuous assurance combines three forms of scrutiny. **Documentary assurance** establishes what was authorised and occurred; **technical assurance** tests data, models, integrations and security; **institutional assurance** examines human reliance, outcomes, complaints, authority and remedies. None suffices alone: accuracy cannot prove legality, while few complaints may signal inaccessible procedure. The useful unit is an *assurance case*: a versioned argument connecting claims about purpose, accuracy, equality, security, reviewability and benefit to controls, evidence, owners, tests and known limitations. NIST connects monitoring with feedback, appeal, override, incidents, recovery, change and decommissioning (Tabassi, 2023, MANAGE 4.1). This is a comparative design resource, not Pakistani law.

5.4.2 The minimum assurance record and monitoring programme

Every consequential system should retain a baseline dossier against which later operation can be compared. It should identify the legal authority and purpose; system owner and accountable official; decision function and affected population; data sources,

provenance, collection dates, transformations, missingness and label definitions; model, dataset, prompt, rule, threshold and software versions; interfaces and downstream recipients; vendor and subcontractors; validation design, subgroup results and known limitations; human-review powers and staffing; security assessment and threat model; complaint route; authorisation conditions; and fallback and exit plan. For each consequential transaction, the deployer should preserve the material input, source-record reference, output or score, model and configuration version, significant factors or explanation, human access and intervention, final action, notice and any correction. Logs must be time-synchronised, access-controlled and tamper-evident, but intelligible to an auditor rather than merely voluminous.

Earlier guidance supports part of this proposal. The NCSC's Guidelines for Secure AI System Development recommend monitoring system behaviour and inputs, with logging that respects privacy and data-protection requirements and supports investigation and remediation (National Cyber Security Centre, 2023). This is comparative guidance, not Pakistani law. Where lawfully operative, the CERT Rules add incident-ticket audit trails and forensic chain-of-custody requirements (*Computer Emergency Response Team Rules*, 2023, rr. 12 and 14). These are cybersecurity controls, not a safe harbour for AI accountability. A retention schedule must reconcile privacy, evidential needs, sector law, appeal and limitation periods; cheap storage does not justify indefinite retention.

Monitoring should track the system's claimed purpose: data completeness and drift; coverage and rejection; calibration, false positives and false negatives; reliability and security; delay; human overrides and automation-biased confirmations; complaints, corrections and appeals; substantive outcomes; and cost and workload against a non-AI baseline. Results should be disaggregated, where lawful and statistically defensible, by groups material to the use. Small samples and missing protected-characteristic data must be reported, not converted into certainty. Sector context matters: calibration may dominate in health, investigative escalation in fraud screening, and language-specific error in translation. Monitoring must follow outputs into linked systems, where one identity error can generate several apparently independent denials.

Testing should be scheduled and trigger-based, with frequency increasing with consequence and volatility. Operators may run routine checks, but high-consequence systems require an assessor independent of development and operations, competent, free of conflicting vendor interests and able to sample real cases and inspect records, logs, versions, incidents and feedback. NIST recognises independent review as a means of reducing internal bias and conflicts of interest and calls for assessors beyond the front-line development team (Tabassi, 2023, MEASURE 1.3). Secure access rooms, redaction, remote testing and protective orders can protect security and trade secrets;

they cannot reduce audit to a vendor-selected demonstration. The public institution must secure sufficient access for a regulator, auditor, ombuds institution or court.

5.4.3 Material change as a reauthorisation trigger

A material change is one that could alter lawful purpose, consequence tier, affected population, performance, security, explanation, human control or remedy. The test is functional, not the vendor's product label. It is triggered by a new data source or linkage; changed schema, imputation, label, sampling or retention practice; retraining, fine-tuning or replacement of a model; changes to prompts, guardrails, retrieval sources, rules or decision thresholds; a new purpose, sector, province, language, population or scale; integration with another database or action system; changed human authority or staffing; a new vendor, owner, subcontractor, cloud location or critical component; a serious vulnerability or changed threat environment; or observed performance, unequal error, complaints or outcomes outside approved limits.

Each deployer should maintain a change register and compare the proposal with the authorised baseline before release. Minor maintenance that cannot plausibly alter output or risk may receive expedited approval, but it must still be logged and tested. A potentially material change should pause production release, reopen the affected parts of the integrated assessment, repeat population-relevant validation and security testing, consult the sector regulator where required, update public information, and obtain reauthorisation from the official who owns the consequence. Emergency security patches may be deployed to contain an active threat, provided that their scope is narrow, the reason and evidence are preserved, safe fallback exists, and retrospective validation follows promptly.

The NCSC's Guidelines for Secure AI System Development recognise that changes to data, models or prompts can change behaviour and recommend treating major updates as new versions with appropriate testing and evaluation (National Cyber Security Centre, 2023). For high-consequence AI, post-change review may be too late where a change can immediately affect liberty, treatment, income or an essential service. The proposed AI protocol should therefore require pre-release review unless emergency containment is necessary. This is a proposed control calibrated to the consequences of change; the international guidance does not itself create a Pakistani statutory approval requirement.

5.4.4 Incident governance: from detection to decommissioning

An AI incident should be defined more broadly than a cyberattack. It includes actual or credibly imminent death or injury; unlawful surveillance or disclosure; wrongful detention, targeting, denial or financial loss; materially discriminatory error; systematic

corruption of source records; unsafe generated content relied upon by a professional; loss of human override; unapproved expansion of purpose or population; serious security compromise; and a near miss revealing that safeguards did not operate. Severity should be assessed from actual and potential harm, scale, sensitivity, reversibility, recurrence, vulnerability of the affected group, critical-service disruption and evidential uncertainty. Notification should not await proof that the model was the sole cause. Early reports may be provisional and updated as causation becomes clearer.

The deploying institution should name a senior incident owner able to preserve evidence, stop automated action, activate continuity and enforce contractual vendor cooperation. Front-line staff and monitors detect and escalate; the data steward preserves lineage; security staff contain compromise; the vendor supplies specified evidence; and a multidisciplinary team investigates data, model, integration, human, procurement, legal and equality causes. Severe or contested cases require independent participation. Attribution must distinguish source-record, transformation, model, threshold, interface, professional and final-action failures. "The algorithm failed" is not a root-cause finding.

Preservation begins immediately. The institution should secure relevant inputs and outputs, model and dataset snapshots, prompts and retrieved material, rules and thresholds, source-record history, access and security logs, human interventions, notices, complaints, contracts and vulnerability reports. Hashing, time synchronisation, access records and an identified custodian protect integrity. Ordinary deletion pauses only for relevant material and as long as lawfully required. Where operative, the CERT Rules treat logs, files and timestamps as incident artefacts and prescribe provable forensic chain of custody (*Computer Emergency Response Team Rules*, 2023, rr. 2, 12 and 14).

Containment must protect people as well as infrastructure. Available measures include disabling automated action while retaining a human-only service; reverting to a validated version; lowering system privileges; isolating a compromised interface; changing a threshold; suspending a data feed; or stopping use for a group in which performance is unsafe. Essential benefits, clinical services and public utilities need a tested manual or alternative channel so that suspension does not itself create mass exclusion. Correction must reach the source record, model or rule that caused the problem and every downstream decision still capable of repair. Retraining is appropriate only if evidence shows that it addresses the cause. Restoration of a benefit, release of a hold, correction of a record and notice to downstream recipients may be more urgent than engineering improvement.

Cyber escalation should use Pakistan's CERT architecture where and to the extent lawfully operative. The published Rules describe national, government, CII, sectoral, federal and provincial routes and, in rule 25, an initial summary to the sectoral and National CERT within one hour of identification, followed by updates and a post-incident report (*Computer Emergency Response Team Rules*, 2023, rr. 3, 8–10,

22–25). Reliance on those provisions requires checking commencement and application to the entity. Institutional incident plans should identify the responsible contacts, evidence to preserve and any additional sectoral or affected-person notification duties. CERT coordination does not displace a sector regulator, PDA within lawful jurisdiction, law enforcement, or affected-person notice, and a CERT does not decide benefit entitlement, negligence or constitutional legality.

The following matrix converts the lifecycle into identifiable duties. Entries marked "proposed" supplement, rather than restate, existing law.

Table 5.4 | Ongoing assurance and escalation

Lifecycle event	Assurance and preserved evidence	Lead and escalation	Transparency and redress
Routine operation	Baseline dossier; versioned performance, subgroup, security, human-reliance and complaint metrics	System owner; independent audit and sector oversight proportionate to tier	Register entry, performance summary and ordinary correction/reconsideration
Material change	Change ticket; differential validation; updated impact, security and fallback assessment	Accountable official; competent regulator or PDA where empowered; CERT where operative; reauthorisation for high consequence (proposed)	Updated notice and consultation where effects materially change
Suspected incident or near miss	Immediate hold on relevant source, model, configuration, output, human-action and security records	Incident owner and multidisciplinary team; vendor duty to cooperate	Interim human review and protective action; preliminary notice where needed
Serious cyber or rights incident	Chain of custody; severity assessment; provisional and updated reports	Sectoral/National CERT under rule 25 if commenced and applicable; sector regulator, PDA, NCHR or law enforcement by subject	Affected-person notice, urgent stay or service restoration; redacted public summary
Recovery and recurrence review	Root-cause analysis; corrections; downstream repair; validated return-to-service decision	Institution retains accountability; independent verification for severe cases	Explanation, restoration/compensation where legally available, audit and policy or contract revision
Uncontrolled or repeated harm	Evidence of failed remediation, opacity, unlawful purpose or unavailable capacity	Authorising body and competent regulator; court/ombuds supervision where invoked	Suspension, restricted use or decommissioning; migration and records plan

Decommissioning is justified when lawful authority has disappeared; performance cannot be validated for the population; serious harm recurs; a vendor prevents necessary audit; security risk cannot be reduced; human review is illusory; or a less intrusive alternative achieves the purpose. Retirement requires a decision record, service-migration plan, return or deletion of vendor-held data, revocation of access, preservation of evidence and notice to regulators and affected users. A system must not be quietly renamed or replaced while its unresolved risk is transferred to a successor.

5.4.5 Consequence-calibrated transparency and participation

Transparency should reveal what people and overseers need without publishing exploit details or other persons' data. At the ordinary tier, a public entry should identify the institution, legal purpose, function, status, responsible contact, broad data categories, human role and complaint route. Enhanced and high-consequence systems should additionally disclose the impact assessment or a reasoned summary; population and geographic scope; significant limitations; validation and disaggregated performance appropriate to the context; audit status; material changes; complaint and correction statistics; and serious-incident summaries. Individual notice and intelligible reasons remain necessary where a person is materially affected. For security-sensitive systems, a public minimum should still identify lawful authority, broad purpose, accountable institution, oversight forum and route to challenge; technical detail may be placed in a confidential annex available to authorised auditors and courts.

This approach fits federal information law. The Right of Access to Information Act 2017 requires federal public bodies to maintain and index records and proactively publish applicable rules, decision processes, criteria, contracts and final performance, audit and investigation reports. Its exemptions protect specified security, privacy, investigatory and commercial interests, but exempt portions must be severed where the remainder can be released; appeal is free and the public body bears the burden of compliance (Right of Access to Information Act, 2017, ss. 4–7, 14, 16–20). Provincial and local bodies fall under provincial regimes. "Vendor confidentiality" is therefore incomplete: the institution must identify the protected interest and disclose non-exempt material. Transparency should not expose biometric templates, personal files, live security configurations or re-identifying data.

Participation is a method of obtaining evidence, not a ceremonial meeting. High-consequence monitoring should provide standing channels through which affected people, front-line officials, unions, professional bodies, disability organisations, women's and minority organisations, researchers and civil society can report unexpected effects. Sampling should reach communities outside major cities. Materials and complaints should be available offline, in Urdu and relevant regional languages, and in accessible formats; travel, connectivity, literacy and retaliation risks matter. Provincial

and local institutions that create records or meet claimants must help shape indicators for devolved services. Subject to lawful limits, consultations should report issues raised, institutional responses and resulting changes. The National AI Policy's multi-stakeholder structure and commitment to citizen engagement, community discussion and public consultation offer a domestic policy foundation (MoITT, 2025, "Policy Implementation Mechanism"). UNESCO supplies comparative human-rights guidance, not a Pakistani cause of action (UNESCO, 2021).

5.4.6 A graduated and connected redress pathway

Redress should begin where the error can be corrected fastest without trapping the person there. The first tier is an accessible institutional channel for source-record correction, preservation of the disputed decision, intelligible reasons and reconsideration by an authorised human able to depart from the output. Urgent cases need a temporary continuation of benefit, pause on enforcement or other protective measure pending review. The channel should be free, usable in person as well as digitally, accessible to persons with disabilities, and capable of accepting a representative's assistance. A single case number should follow the complaint, and a "wrong door" should result in assisted referral rather than loss of a limitation period. Reconsideration outcomes and reasons should feed back into monitoring so that repeated "individual" errors are recognised as systemic.

The second tier is the competent sector or accountability forum, not a generic AI office. Depending on subject and jurisdiction, that may be a line department's statutory appeal body; a regulator in banking, securities, telecommunications, energy, health or another regulated field; a specialised tribunal; a professional council; a federal or provincial ombuds institution; an information commission; a rights commission; or an audit body. Their functions are distinct. A sector regulator can test licence, service and professional obligations; a professional body can address unsafe reliance; an information commission can order access to non-exempt records; the Auditor-General can examine expenditure, procurement and control failure but ordinarily does not restore an individual benefit; and NCHR may inquire on petition or *suo motu* into human-rights violations or public-servant neglect and recommend measures within its statutory mandate (National Commission for Human Rights Act, 2012, ss. 9 and 18). Jurisdictional limits, including special security and intelligence provisions, must be acknowledged rather than obscured by an AI label.

Ombuds institutions are particularly useful for maladministration that combines poor records, delay, arbitrary reliance and failure to hear a person. Under the Federal Ombudsmen Institutional Reforms Act 2013, the covered ombudsman has civil-court powers for temporary injunction and implementation, may stay an impugned order for up to sixty days, and has review and presidential-representation procedures (Federal

Ombudsmen Institutional Reforms Act, 2013, ss. 9–14). The exact ombudsman general, tax, banking, insurance, workplace-harassment or provincial and any exclusion depend on the governing enactment. It would be misleading to present ombuds review as universally available for every AI-assisted decision.

The final tier includes ordinary civil or criminal proceedings where their elements are satisfied and constitutional judicial review. High Courts exercise Article 199 jurisdiction through Constitutional Benches where Article 202A has been brought into force; following the Twenty-Seventh Amendment, the Federal Constitutional Court has the specified public-importance fundamental-rights jurisdiction and hears qualifying appeals from Article 199 orders (Constitution of the Islamic Republic of Pakistan, 1973, arts. 175E–175F and 199, as amended in 2025). Judicial review should remain available against the public authority's decision even when the model, cloud or evidence is vendor-controlled. Interim relief is essential where detention, treatment, subsistence, demolition, blacklisting or an irreversible disclosure is imminent.

Remedy must match urgency, reversibility, scale and recurrence. Individual measures include correction, explanation, fresh human decision, restoration of benefit or status, deletion or restriction of unlawfully used data, notification of downstream recipients, injunction and compensation where a legal basis exists. Systemic measures include independent audit, expanded sampling, threshold change, revalidation or retraining, revised forms and source records, staff instruction, contract enforcement, procurement revision, incident notification, regulator reporting, continuing supervision, suspension and decommissioning. A court or regulator should not order retraining merely because it sounds technical: if the defect is unlawful purpose, missing authority, inaccessible appeal or structurally incomplete records, a more accurate model would reproduce the legal failure. Conversely, deciding one claimant's case without correcting a shared data pipeline leaves foreseeable harm in place. Effective redress closes the assurance loop by converting reasons, complaints, incidents and remedies into evidence for continued authorisation.

5.4.7 Conclusion

Ongoing assurance is neither permanent surveillance of every low-impact tool nor an annual certificate filed and forgotten. It is a consequence-calibrated system of traceable evidence, independent access, material-change control, rights-sensitive incident response, proportionate disclosure, participation and connected redress. Pakistan can build much of it through instruments already in place, while reserving coercive duties, new liabilities and jurisdictional expansion for clearly authorised legal forms. The decisive principle is continuity of responsibility: deployment never converts an institution's public or professional duty into a vendor's technical problem.

5.5 Phased Implementation, Capacity, Cost, and Institutional Evaluation

Implementation should begin with a proposition that is modest but demanding: Pakistan need not wait for a comprehensive artificial-intelligence Act before improving the governance of systems already being considered, bought or used, but neither may an administrative checklist create powers that the Constitution or legislation has not conferred. The sequence must therefore be **authority-sensitive rather than technology-led**. Measures available under procurement, financial-management, records, cybersecurity, professional and sector mandates should start now; intergovernmental measures should proceed through lawful federal and provincial channels; and primary legislation should be reserved for duties, rights and coercive powers that existing law cannot supply. The objective is not an appearance of readiness created by a policy, committee or training event. It is demonstrable institutional ability to authorise, supervise, challenge, finance and, where necessary, stop a consequential system.

The *National Artificial Intelligence Policy 2025* is an approved executive policy. The *Digital Nation Pakistan Act 2025* establishes the National Digital Commission and Pakistan Digital Authority (PDA), and gives them enacted functions concerning Masterplan directives and compliance, standards, and oversight of assigned artificial-intelligence and data-governance matters. Those powers remain subject to the Act's requirement that PDA avoid overlap with existing regulators and conflict with existing law, and that designated custodians retain control and responsibility for their data (*Digital Nation Pakistan Act*, 2025, ss. 5, 8(b), 8(g)–(h), 12; Ministry of Information Technology and Telecommunication [MoITT], 2025). These are relevant foundations, but they do not convert every policy safeguard into an enforceable right or transfer all constitutional, sectoral and provincial competence to one body.

Sequencing instruments by legal authority

The first track is an **immediate administrative baseline**, capable of being introduced within existing authority and without claiming a new general AI jurisdiction. Each federal division, provincial department, regulator, public entity and local institution should identify systems that classify, rank, predict, match, recommend, generate, detect anomalies or automate action; record the documentary status used in Chapter 4; assign a senior responsible official; and identify the legal provision, decision function, data custodian, supplier, affected population, consequence tier, complaint route and current contract. This inventory is a factual and managerial exercise. It does not authorise a system merely by listing it. An institution unable to establish authority for a consequential use should suspend acquisition or expansion while obtaining legal advice, and cease the use if no valid basis exists.

Administrative directions and standard operating procedures can then require documented case reasons, source-record correction, human escalation, version and decision logging, access controls, preservation schedules, and approval for material change. At federal level such directions must follow the Constitution, the *Rules of Business 1973* and the enabling law of the relevant body; provinces and autonomous institutions require their own lawful instruments (Government of Pakistan, Cabinet Division, 2025). These directions govern officials and institutional workflow. They cannot create offences, impose penalties on an unregulated private party, compel a province, displace a statutory appeal, or legalise surveillance or data linkage that lacks substantive authority.

The second track is **procurement and public-finance reform**. The competent federal or provincial procurement authority and, where its enactment requires, the relevant government may develop AI and algorithmic-system schedules, templates or guidance only within that regime's rule-making and approval structure. Pakistan has distinct federal, Punjab, Sindh, Khyber Pakhtunkhwa and Balochistan regimes; a federal template is consequently a model for adaptation, not a nationwide command. Existing procurement principles concerning planning, competition, evaluation, transparency and value for money provide an institutional entry point for requirements on provenance, validation, audit access, subcontractors, security, portability, incidents and exit (*Public Procurement Rules*, 2004; *Punjab Procurement Rules*, 2014; *Sindh Public Procurement Rules*, 2010; *Khyber Pakhtunkhwa Public Procurement Regulatory Authority Act*, 2012; *Balochistan Public Procurement Regulatory Authority Act*, 2009).

The practical reform is to make lifecycle evidence part of responsiveness and contract performance, rather than an ethical appendix examined after award. A procuring agency should state the lawful problem and baseline before specifying a product; compare AI with record repair, staffing or simpler software; require evidence proportionate to the consequence tier; evaluate whole-life cost; and refuse terms that prevent lawful audit, complaint resolution or migration. Framework contracts can supply common testing or specialist services where the applicable rules permit, but should not force every department into one model or vendor. The World Bank's GovTech procurement guidance similarly treats total cost of ownership, interoperability, contract management and disposal as procurement questions, while OECD analysis identifies data and vendor lock-in as public-sector AI risks (OECD, 2025; World Bank, 2021).

Financial controls should operate alongside tender controls. Federal projects can incorporate governance work packages, recurrent expenditure, measurable outputs and closure obligations into the applicable budget and development-project processes. The *Public Finance Management Act 2019* requires performance-oriented budgeting and year-end reporting on expenditure by outputs and delivered performance targets; it provides a basis for asking what a system achieved with public funds, though not for

inventing an AI-specific legal power (*Public Finance Management Act*, 2019, ss. 9, 35). Provincial finance and planning systems should make equivalent provision through their own law and procedures. No high-consequence project should receive approval on a capital estimate that omits data repair, professional oversight, independent testing, complaints, security operations and exit.

The third track is **sectoral rule-making and supervised pilots**. A banking, securities, telecommunications, health, education, taxation, policing or employment use should be governed by the competent body or bodies at each lifecycle stage, applying their distinct service, professional, territorial and remedial mandates while coordinating with procurement, cybersecurity, audit, data-governance and rights institutions. Where an enabling statute permits regulations, licence conditions, codes or directions, the competent body can translate the common protocol into sector-specific validation, disclosure, human-review and incident requirements. A clinical support system, for example, needs professional validation and escalation different from a tax anomaly tool. A pilot should begin only after these conditions and stop rules exist; "pilot" must not become a label for operational exposure without authority or remedy.

Capacity-building is a fourth, continuous track, but training must follow assigned duties. Introductory AI literacy is insufficient for an official expected to test calibration, a procurement officer expected to negotiate evidence rights, a doctor expected to interpret uncertainty, or an ombuds investigator expected to reconstruct a decision chain. A Pakistan-specific study illustrates the point without establishing national prevalence. Interviews with representatives of 27 university libraries found adoption at a nascent stage, shaped by technical practice, finance, organisational size, data management and protection concerns (Jan et al., 2024). Readiness is therefore institution- and function-specific, not inferred from a national policy or a general course certificate.

Primary legislation forms the fifth track, proceeding in parallel where delay would leave an unlawful or unremedied power gap. Legislation is genuinely necessary where the competent federal or provincial legislature chooses, within its constitutional field, to create enforceable rights concerning consequential automated decisions; regulate personal-data processing; prohibit defined practices; authorise intrusive powers with safeguards; compel evidence from actors outside existing jurisdiction; create penalties or inspection powers; establish a new appellate route; or give a coordinating body coercive authority beyond its present mandate. Sector statutes may instead need targeted amendment. The case for legislation should be supported by mapped gaps and pilot or enforcement evidence, not by the assumption that every difficulty requires one omnibus Act or one AI authority. Conversely, a pilot cannot be used to test whether government may exercise a coercive power that applicable law has not granted.

The following matrix combines the implementation sequence with evidence of readiness. Time periods are proposed planning ranges, not existing legal deadlines, and may overlap where authority and capacity permit.

Table 5.5 | Phased implementation and stopping conditions

Proposed phase	Principal instruments and actions	Minimum readiness evidence before progression	Hold, restrict or stop if
0–6 months: establish control	System inventory; documentary status; named owner; interim legal, data, records and security check; freeze hidden expansion; map complaints and contracts	Identified authority, decision chain, data custodian, vendor, affected population, current use and preservation route	Authority, owner, operative version or consequential use cannot be established
6–18 months: build common foundations	Procurement schedules and model clauses; register and assessment templates; budget tags; sector and provincial adaptation; role-based training; audit and institutional incident-routing protocols, including CERT routes where lawfully operative	Trained multidisciplinary team, records and test environment, funded complaint and incident routes, supplier access commitments	Safeguards exist only on paper, recurrent funding is absent, or vendor access is refused
12–30 months: controlled authorisation	Consequence-classified pilots; local and subgroup validation; independent evaluation; sector conditions; public or protected reporting	Baseline, thresholds, human fallback, accessibility, security tests, independent evaluator and exit resources	Stop threshold is crossed, benefit is unproved, unequal harm persists, or service continuity fails
24–48 months: targeted scaling and legislation	Scale only successful uses; amend sector statutes; enact cross-sector duties or powers where gap analysis demonstrates need; formalise intergovernmental arrangements	Published evaluation, legislative competence analysis, implementation budget, regulator and judicial capacity, provincial consultation where relevant	Legislation assigns unfunded or duplicative functions, obscures remedies, or exceeds competence
From 36 months: consolidation and renewal	Reauthorisation, portfolio review, audit, contract renewal or re-procurement, standards and statutory review, decommissioning	Comparable outcome and cost data, complaint and incident history, continuing authority, current validation and funded exit	Objectives are unmet, evidence is unavailable, residual harm is disproportionate, or a safer alternative performs better

Readiness as a non-compensable authorisation gate

A readiness assessment should serve two different questions: can the institution **deploy** this system, and can the responsible public body or regulator **supervise** it? A supplier's technical capacity answers neither. For deployment, the institution needs a valid legal basis and competent decision-maker; defined purpose and prohibited uses; a data custodian able to attest provenance, quality, linkage and correction; validation against

the actual population, languages and operating conditions; secure infrastructure and business continuity; meaningful human review; accessible notice, complaint and urgent correction; reproducible logs and preservation; contractual access to the supplier and subcontractor chain; a funded monitoring and exit plan; and documented coordination where records or services cross governmental boundaries.

Supervisory readiness requires an enabling mandate, operational independence appropriate to the function, investigators and domain experts, protected access to data and vendor records, secure facilities for examining confidential material, power to obtain or lawfully request information, complaint triage, interim and final remedies, audit coordination, incident referral and a reasoned enforcement policy. Courts, tribunals and ombuds institutions do not need to become model developers, but they need sufficient technical literacy, procedural tools and independent expertise to test provenance, reliability and institutional reliance. An organisation should not be designated the supervisor merely because it hosts a digital programme.

Readiness may be displayed as a dashboard, but certain conditions are **non-compensable**. No aggregate score can offset absent legal authority, inability to preserve evidence, lack of an effective correction or review route, a critical unresolved vulnerability, or inability to maintain the underlying public service when the system is withdrawn. Other weaknesses may justify a bounded pilot or a lower scale: limited compute, for example, may be addressed through a controlled environment; weak language validation requires restriction to validated users and purposes. Every claimed capability should be evidenced by a named person, document, test, budget line or exercised procedure, with an expiry or reassessment date.

The assessment should also test actual workload. A nominal human reviewer may be responsible for thousands of alerts, an internal lawyer may lack access to the model contract, or a complaint office may be available only through the digital channel that produced the exclusion. Simulation and tabletop exercises should therefore precede high-consequence deployment: staff should reconstruct a sample decision, correct a source record, handle an override, respond to a security incident, preserve evidence, notify an affected person and switch to the fallback. Failure in rehearsal is evidence of institutional unreadiness, not a training footnote. NIST's voluntary AI Risk Management Framework usefully links governance, mapping, measurement and management across the lifecycle. UNESCO's Ethical Impact Assessment supplies questions about the purpose, project team, stakeholder participation and safeguards of a proposed system, including whether automation is appropriate in the first place (Tabassi, 2023; UNESCO, 2023). These are design resources, not sources of domestic power.

Costing credible lifecycle oversight

Acquisition price is a poor estimate of public cost. Each business case should include at least four cost accounts. The **technical account** covers problem definition, data inventory and cleaning, labelling, integration, compute, storage, connectivity, licences, localisation, testing, cybersecurity, updates, drift monitoring, redundancy and service continuity. The **governance account** covers legal analysis, integrated impact assessment, register maintenance, documentation, independent validation and audit, secure disclosure, incident investigation, records preservation, regulator access and public reporting. The **people and remedy account** covers professional supervision, reduced caseloads where genuine review requires time, role-based training, accessible and offline notice, translation, disability accommodation, complaint staff, legal assistance, expert support, reconsideration, restoration and compensation. The **exit account** covers parallel or manual fallback, data and model portability, supplier transition, contract termination, archival preservation, secure return or deletion, downstream record correction, staff redeployment and decommissioning verification.

Budgeting must distinguish initial, recurrent, contingent and exit expenditure. Monitoring, licences and skilled staff are recurrent; breach response, urgent expert inspection and compensation are contingent but foreseeable; migration and deletion are eventual rather than optional. Funds for these functions should be identified before authorisation and protected against the common pattern in which a development budget purchases a platform while the operating institution inherits unfunded obligations. The federal performance-budget framework already links allocations with outputs and outcomes; AI projects should use that machinery to disclose the combined recurrent and development cost rather than claim savings against only the purchase price (Government of Pakistan, Finance Division, 2022).

The proper comparison is the system's discounted whole-life cost against both the existing process and feasible alternatives, including record improvement, simpler rules, additional staff or a modular public digital service. Costs shifted to another institution or to an affected person remain social and administrative costs: repeated travel to correct a record, lost wages while appealing, additional clinician verification, litigation, or local-office workload should not be booked as central "efficiency". Sensitivity analysis should model exchange-rate exposure, usage-based cloud and application-programming-interface charges, model replacement, supplier failure and a sharp increase in complaints or incidents. Benefits should likewise be evidenced rather than promotional: reduced processing time is not a saving if error, exclusion or later review increases.

Shared capacity can reduce duplication. A federal or provincial centre may maintain model clauses, approved testing methods, secure expert panels and training modules, while universities and accredited laboratories may support independent evaluation. Yet

pooled services need conflict-of-interest controls and cannot replace the accountable sector decision. Smaller provinces, districts and local bodies require funded implementation support rather than identical unfunded duties. Cost proportionality should simplify documentation for ordinary uses; it should not waive minimum authority, security, records or remedy for a high-consequence use because the institution is under-resourced. Where credible oversight is unaffordable, the lawful answer may be not to deploy.

Performance, distribution and institutional evaluation

Evaluation begins before procurement with a stated lawful objective, baseline, feasible comparator, indicators, data-collection plan and decision thresholds. Vendor accuracy on a benchmark is not proof that a deployed system improves public administration. Where practicable, a pilot should compare outcomes with the prior process or a credible non-AI alternative and separate the effect of the model from contemporaneous changes in staffing, records or procedure. Quantitative estimates should report uncertainty, missingness and sample limitations; qualitative evidence from officials, professionals and affected groups should identify harms or burdens that an aggregate metric misses.

Five linked dimensions determine continuation. **Substantive performance** asks whether the lawful objective improved: not merely how often a prediction matched a label, but whether eligible people received benefits more accurately, inspections found legally relevant non-compliance, clinicians were assisted safely, complaints were resolved, or processing time fell without displaced error. It includes false-positive and false-negative rates, calibration, coverage, reliability, drift and human override. The threshold must reflect the consequence of each error; a single average "accuracy" figure is inadequate.

Distributional performance disaggregates error, delay, non-coverage and benefit by legally and contextually relevant groups, including gender, disability, age where relevant, language, province and district, rural residence, income, migration or documentation status and digital access. Small samples require careful interpretation rather than suppressed concern. Evaluation should examine intersectional and cumulative effects, source-record correction, and whether errors propagate into linked systems. Equality is not established by equal average treatment if one group is systematically missing from the data or bears a more severe consequence.

Security and resilience measure tested vulnerabilities, access violations, poisoning or evasion attempts, dependency and patch status, time to detect, contain and recover, integrity of backups and logs, subcontractor incidents, service availability and success of fallback exercises. The approved *National Cyber Security Policy 2021* supplies policy context, while the published *Computer Emergency Response Team Rules, 2023* design an incident-coordination structure, including sectoral post-incident reporting according to

severity, impact, scale and resolution. Rule 1(3) of the CERT Rules, however, makes commencement dependent upon a further Gazette notification, so application requires confirmation of commencement and the relevant mandate. AI-specific evidence should connect to that structure where and to the extent lawfully operative, while procurement and institutional incident plans should not wait for uncertainty to be resolved (*Computer Emergency Response Team Rules*, 2023; MoITT, 2021). Absence of reported incidents is not proof of security if detection, reporting or independent access is weak.

Procedural and remedial performance measures whether notice identifies automation and the responsible institution; reasons expose material data and factors; reviewers have time and authority to depart; source records can be corrected; urgent cases receive interim protection; and decisions, overrides and changes are reproducible. Indicators include notice delivery, explanation adequacy, override and disagreement patterns, correction and reconsideration times, appeal outcomes, restoration, repeat failure and record completeness. Complaint volume must be normalised by exposure and interpreted with accessibility evidence. Very few complaints may indicate satisfaction, or that people do not know about the system, cannot reach the forum, fear retaliation or cannot use a digital portal.

Institutional and financial performance measures staff workload, automation bias, vacancies, training demonstrated in practice, supplier responsiveness, audit findings, coordination failures, recurrent and unit cost, cost of error and remedy, and total cost against alternatives. The Auditor-General's constitutional and statutory financial and performance-audit functions extend through the relevant provincial audit directorates, with reports proceeding to the competent legislature and Public Accounts Committee; an algorithmic audit protocol should add data, model, security and rights expertise without narrowing that mandate (*Constitution of the Islamic Republic of Pakistan*, 1973, arts. 168–171; *Auditor-General's (Functions, Powers and Terms and Conditions of Service) Ordinance*, 2001; Auditor-General of Pakistan, n.d.). Sector regulators, internal audit, ombuds institutions and other bodies retain their own lawful roles.

Results should be recorded in a review dossier linking each indicator to evidence, threshold, responsible actor and corrective action. It should state whether the system will continue unchanged, continue with conditions, be narrowed, retrained, re-procured, suspended or terminated. Public reporting may aggregate or redact genuinely sensitive material, but a confidential annex must preserve enough information for the competent regulator, auditor or court. Failure to produce the operative version, validation set, decision logs or contractually controlled records is itself an assurance failure; uncertainty created by missing evidence should not be treated as successful performance.

Periodic review, suspension and termination

Scheduled review complements the material-change and incident triggers in Section 5.4. A newly authorised high-consequence system should receive an early post-deployment review, regular internal dashboards and independent review on a risk-proportionate cycle fixed in the authorisation. Rapidly changing, security-exposed or difficult-to-reverse uses require shorter cycles; a stable use may justify a longer interval only after evidence accumulates. Authorisations and pilots should expire unless affirmatively renewed. Sector regulations and common standards should state their own review dates, while primary legislation establishing novel AI duties should require an early published implementation review and periodic parliamentary or provincial legislative scrutiny thereafter. These are design requirements, not claims about existing Pakistani statutory periods.

Review must test both the system and the supervising institution. Revision is appropriate where a remediable threshold, interface, notice, training or contract term is defective. Restriction may narrow a population, function, data source or consequence. Suspension is required where a serious security or rights incident is under investigation, validation no longer matches the population, lawful authority is disputed, material records are missing, human fallback has failed, the supplier denies required access, or credible evidence indicates disproportionate group harm. Urgent suspension should preserve essential service through the pre-funded fallback and prevent the automated output from continuing to generate adverse effects.

Termination is justified when the lawful objective is not achieved; comparative performance or whole-life cost is inferior to a feasible alternative; recurrent harm remains disproportionate; security cannot be restored; responsibility or authority cannot be made clear; or the institution cannot sustain oversight. It is not enough to cancel a licence. Decommissioning must stop new outputs, preserve evidence for pending claims, notify responsible forums and affected persons where appropriate, restore or reconsider decisions, correct source and downstream records, return or securely delete data according to law, remove integrations and credentials, verify supplier compliance, maintain service continuity, and publish a closure evaluation. Contract renewal should never operate as tacit reauthorisation.

The staged package is therefore concrete: inventory and stabilise; reform procurement and budgeting; build role-specific capacity and auditable records; permit only bounded, evaluated pilots; adopt sector rules within lawful mandates; legislate targeted gaps; measure outcomes and distribution over the whole lifecycle; and renew only on affirmative evidence. Pakistan's uneven capacity is a reason to sequence, support and sometimes decline deployment not to dilute rights, security or review into aspirations that no institution can perform.

Chapter 5 Key Arguments

Regulatory scope must follow function and consequence, not architecture or marketing. Prediction, classification, matching, ranking, recommendation, generation, anomaly detection and automated action become consequential when they structure public authority, security, healthcare, education, work, finance or services. A rules engine may therefore demand stronger oversight than a sophisticated model used for minor assistance. Classification turns on affected interests and populations, scale, data sensitivity, reversibility, practical human control, security exposure and cumulative effect. It must remain revisable when purpose, population, data, model, threshold, integration, vendor, threat environment or observed outcome changes. Missing evidence is not proof of low risk.

Pakistan should coordinate existing institutions rather than presume that one new authority must displace them. National coordination can supply terminology, registers, assessment templates, procurement clauses, incident formats, expertise, intergovernmental procedure and reporting; it must not assume powers assigned elsewhere. Sector regulators and departments retain the context needed to judge medicine, banking, work, education, policing, benefits and municipal delivery. Provinces and local bodies must participate where they create records, deliver devolved services or receive complaints. Cybersecurity, procurement, audit, ombuds, rights, professional and judicial bodies perform distinct functions and should remain accessible. The model is hybrid in operation but source-specific in law: each duty and remedy requires a competent body and instrument.

High-consequence use requires controls before public exposure. Before procurement becomes commitment, the institution should record lawful purpose and authority, responsible officials, population, data, system and vendor, human role, validation, security, complaint route and exit. An integrated assessment must join authority with provenance, equality, accessibility, cybersecurity, model integrity, professional practice, federal fit, cost and remedy. Approval should be withheld where relevant populations and languages remain unvalidated, logs cannot reconstruct an output, humans cannot depart, the supplier blocks audit or evidence access, continuity is unfunded, or no lawful fallback exists. Procurement must preserve these capabilities through subcontracting, updates, incidents and termination; contractual risk allocation does not extinguish public or statutory duties.

Authorisation begins rather than completes governance. Assurance requires versioned records, traceability, justified independent access, monitoring of performance and unequal error, human-reliance and complaint analysis, security testing and confirmation of continuing lawful purpose. Material changes require reassessment, not concealment as maintenance. Before failure, incident plans must allocate detection, preservation, investigation, containment, notification, suspension, correction, continuity and

decommissioning. Transparency should be calibrated: legitimate confidentiality may control access to sensitive material but cannot justify missing records, decision-makers or review. Affected communities, professionals and independent researchers should inform evaluation where they can reveal burdens hidden by aggregate accuracy.

Redress must join technical correction to legal and institutional remedy. A person may need source-record correction, reasons, reconsideration, restored benefit, interim relief or compensation. Recurrent harm may also require affected-population notice, independent audit, new thresholds, data repair, revalidation, contract enforcement, procurement or policy revision, suspension or decommissioning. Internal review should connect without becoming a dead end to the competent regulator, ombuds institution, tribunal, commission, auditor, professional body or court. The AI system is never the legal actor. Responsibility follows control and duty across officials, professionals, custodians, developers, integrators, cybersecurity providers, vendors and deployers, while the public body retains responsibility for its authority.

Implementation must be staged, costed and capable of retreat. Existing powers can support inventories, procurement specifications, logging standards, administrative controls, sector guidance, audits, training and bounded pilots. Delegated rules or licence conditions require enabling authority; primary legislation is necessary for new coercive powers, cross-sector duties, offences, jurisdiction or generally enforceable rights. Readiness must test staffing, data stewardship, security, audit, complaints, legal expertise, provincial coordination, independent evaluation and budget not merely acquisition. Lifecycle cost includes assessment, validation, monitoring, incidents, preservation, legal assistance, reporting and exit. Periodic review should compare lawful objectives, errors, distributional effects, security, workload, complaints, reviewability and total cost against realistic alternatives. Inadequate evidence, disproportionate harm, unclear responsibility or false capacity assumptions justify revision, suspension or termination.

References

Aas, K., Jullum, M., & Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence, 298*, 103502. https://doi.org/10.1016/j.artint.2021.103502

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 308–318). Association for Computing Machinery. https://doi.org/10.1145/2976749.2978318

Abbas, A. (2024). Urban Pakistan and women's mobility in securitised Lahore. *Gender & Development, 32*(1–2), 511–521. https://doi.org/10.1080/13552074.2024.2348385

Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2018). Sanity checks for saliency maps. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 31* (pp. 9505–9515). Curran Associates. https://proceedings.neurips.cc/paper/8160-sanity-checks-for-saliency-maps

Ahmad, Z., Shah, I., Alam, A., & Khan, H. (2025). Quality education in post-devolution scenario in Pakistan: Evidence from Khyber Pakhtunkhwa province. Development in Practice. Advance online publication. https://doi.org/10.1080/09614524.2025.2558748

Ahmed, I., Mustafa, U., & Khalid, M. (2007). National Finance Commission awards in Pakistan: A historical perspective (PIDE Working Papers 2007:33). Pakistan Institute of Development Economics. https://pide.org.pk/research/national-finance-commission-awards-in-pakistan-a-historical-perspective/

Ahmed, S., Yaakub, A. N., & Javed, A. (2025). Artificial intelligence in Pakistan's cyberspace: Governance gaps, dual-use dilemmas, and strategic vulnerabilities. Obrana a strategie, 25(2), 3–23. https://doi.org/10.3849/1802-7199.25.2025.02.3-23

American Bar Association Standing Committee on Ethics and Professional Responsibility. (2024, July 29). *Formal Opinion 512: Generative artificial intelligence tools* [Comparative professional guidance; not Pakistani law]. https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf

Ara, I., & Sabir, M. (2010). Decentralisation of GST services and vertical imbalances in Pakistan. The Pakistan Development Review, 49(4, Part II), 479–495. https://doi.org/10.30541/v49i4IIpp.479-495

Arif, S., Azeemi, A. H., Raza, A. A., & Athar, A. (2024). Generalists vs. specialists: Evaluating large language models for Urdu. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 7263–7280). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.426

Askari Bank Ltd. v A.H. International (Pvt) Ltd. and others, 2025 SHC KHI 2590 (Sindh High Court, Execution No. 25 of 2012, order of 23 September 2025). https://caselaw.shc.gov.pk/caselaw/view-file/MjgzNTczY2Ztcy1kYzgz

Auditor-General of Pakistan. (2012). *Financial Audit Manual* (Rev. ed.). https://agp.gov.pk/SiteImage/Misc/files/Revised%20FAM%20Book%281%29.pdf

Auditor-General of Pakistan. (n.d.). *Performance Audit Wing* [Official institutional description; not evidence of specialist AI-audit capacity]. https://www.agp.gov.pk/Detail/MjE4MWE0OTYtMzQzNC00MDVlLWIyYjEtMzY0MDVkY2YyN2Uw

Auditor-General's (Functions, Powers and Terms and Conditions of Service) Ordinance, 2001 (Ordinance XXIII of 2001; 17 May 2001; commenced 1 July 2001), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/pdffiles/administratord2a43e2e40d8af8e23fd95cba332137b.pdf

Axelsson, S. (2000). The base-rate fallacy and the difficulty of intrusion detection. *ACM Transactions on Information and System Security, 3*(3), 186–205. https://doi.org/10.1145/357830.357849

Aziz, S. (2018). The constitution of Pakistan: A contextual analysis. Hart Publishing. https://doi.org/10.5040/9781509995424

Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). How to backdoor federated learning. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (Proceedings of Machine Learning Research, Vol. 108, pp. 2938–2948). PMLR. https://proceedings.mlr.press/v108/bagdasaryan20a.html

Balochistan Public Procurement Regulatory Authority Act, 2009 (Balochistan Act VIII of 2009), as amended/current at the cut-off. Balochistan Code, official text. https://balochistancode.gob.pk/Document.aspx?docc=1918&docid=2053&wise=opendoc

Banking Companies Ordinance, 1962 (Ordinance LVII of 1962), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/index.php/UY2FqaJw1-apaUY2Fqa-cJ%2BV-sg-jjjjjjjjjjj

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732. https://doi.org/10.15779/Z38BG31

Benazir Income Support Programme Act, 2010 (Act XVIII of 2010; 18 August 2010), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2Npa5pjZg%3D%3D-sg-jjjjjjjjjjj

Benazir Income Support Programme. (n.d.-a). *8171 web portal* [Operational public query channel; not proof of merits review in each case]. https://8171.bisp.gov.pk/

Benazir Income Support Programme. (n.d.-b). *Benazir Kafaalat programme* [Official programme description]. https://bisp.gov.pk/Detail/YTgzNjkxM2YtN2ViMC00MjA5LWI0MDMtNzM4ZWJmMGVlNzc5

Benazir Income Support Programme. (n.d.-c). *Benazir National Socio-Economic Registry (NSER)* [Official registry description; not evidence of a neural-network eligibility model]. https://bisp.gov.pk/Detail/NzI5YTMyYTMtYjE1My00NGUwLTgwYTItZWUwYTZkYWZjYmNj

Benazir Income Support Programme. (n.d.-d). *BISP–NADRA sign MoU for new survey under Dynamic Registry* [Official institutional announcement]. https://bisp.gov.pk/NewsDetail/MGJjYmY5YjAtM2IwZS00ZTVmLWE5ZGQtNWM0MzNmMWNlZTlm

Benazir Income Support Programme. (n.d.-e). *Complaint redressal* [Official complaint route; existence does not establish accessibility or outcome in every case]. https://bisp.gov.pk/Detail/ZDk4ZWViMTUtZWQ4ZS00ZjM4LTk3ZGYtYzlkOTAwMGVkOWY4

Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. *Information, 10*(4), Article 122. https://doi.org/10.3390/info10040122

Binns, R. (2022). Human judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance, 16*(1), 197–211. https://doi.org/10.1111/rego.12358

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1175–1191). Association for Computing Machinery. https://doi.org/10.1145/3133956.3133982

Buiten, M. C., de Streel, A., & Peitz, M. (2023). The law and economics of AI liability. *Computer Law & Security Review, 48*, Article 105794. https://doi.org/10.1016/j.clsr.2023.105794

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (Proceedings of Machine Learning Research, Vol. 81, pp. 77–91). PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html

Butt, S. A., Pappel, I., & Oolu, K. (2021). Implementation of electronic records management systems: Potential and challenges a case study of the Water and Power Development Authority (WAPDA) in Pakistan. In S. Fong, N. Dey, & A. Joshi (Eds.), ICT analysis and applications: Proceedings of ICT4SD 2020, volume 2 (Lecture Notes in Networks and Systems, Vol. 154, pp. 629–639). Springer Singapore. https://doi.org/10.1007/978-981-15-8354-4_63

Carlini, N., Tramèr, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, Ú., Oprea, A., & Raffel, C. (2021). Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)* (pp. 2633–2650). USENIX Association. https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting

Centre for Labour Research. (n.d.). *Fairwork Pakistan* [Project page]. https://clr.org.pk/fairwork-pakistan/

Chandra, B., Dunietz, J., Roberts, K., Lee, Y., Fontana, P., & Awad, G. (2024). *Reducing risks posed by synthetic content: An overview of technical approaches to digital content transparency* (NIST AI 100-4) [Comparative technical guidance; not Pakistani evidence law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-4

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data, 5*(2), 153–163. https://doi.org/10.1089/big.2016.0047

Clark, J. M., Haseeb, M., Jalal, U. E. A., Siddiqi, B. M., & Vyborny, K. H. (2022). *Using biometrics to deliver cash payments to women: Early results from an impact evaluation in Pakistan* (ID4D Evidence Note). World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099155004142238180

Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 598–609). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445921

Code of Civil Procedure, 1908 (Act V of 1908), as amended/current at the cut-off, especially s. 9 and Orders XI and XXXIX. Pakistan Code, official text. https://pakistancode.gov.pk/pdffiles/administrator6598dabbad120033d4d42d717dcf9755.pdf

Coglianese, C. (2024). Procurement and artificial intelligence. In R. Paul, E. Carmel, & J. Cobbe (Eds.), *Handbook on public policy and artificial intelligence* (pp. 235–248). Edward Elgar Publishing. https://doi.org/10.4337/9781803922171.00026

Companies Act, 2017, as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2Npa5Zm-sg-jjjjjjjjjjj

Competition Act, 2010 (Act XIX of 2010), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2FsbJk%3D-sg-jjjjjjjjjjj

Computer Emergency Response Team Rules, 2023 (S.R.O. 1394(I)/2023; notification dated 26 September 2023; Gazette publication 3 October 2023) [Rule 1(3) requires a further commencement notification; applicability requires confirmation]. National CERT, official Gazette copy. https://pkcert.gov.pk/uploads/2023/10/GAZETTE-CERT-Rules-2023.pdf

Constitution of the Islamic Republic of Pakistan, 1973, as amended through 21 November 2025, incorporating the Constitution (Twenty-seventh Amendment) Act, 2025. National Assembly of Pakistan, official consolidated text. https://na.gov.pk/uploads/documents/6926e060076ed_467.pdf (Official searchable Pakistan Code text: https://pakistancode.gov.pk/pdffiles/administrator9d8e2ecc414c6d3371ac41114b61a2c4.pdf)

Constitution (Twenty-seventh Amendment) Act, 2025 (Act XXXII of 2025), Gazette of Pakistan, Extraordinary, 13 November 2025. National Assembly of Pakistan, official text. https://www.na.gov.pk/uploads/documents/691ec19a6a212_270.pdf

Contract Act, 1872 (Act IX of 1872), as amended/current at the cut-off, especially ss. 73–74. Pakistan Code, official text. https://pakistancode.gov.pk/pdffiles/administrator8332a6df32386960ac7d81a5cf7aade2.pdf

Cotino Hueso, L. (2023). Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida [What concrete transparency and information from algorithms and artificial intelligence is due]. *Revista Española de la Transparencia, 16*, 17–63. https://doi.org/10.51915/ret.272

Crawford, K., & Schultz, J. (2019). AI systems as state actors. *Columbia Law Review, 119*(7), 1941–1972. https://columbialawreview.org/content/ai-systems-as-state-actors/

Customs Act, 1969 (Act IV of 1969), as amended up to 30 June 2025. Federal Board of Revenue, official consolidated text. https://download1.fbr.gov.pk/Docs/20258121285942396CustomsAct1969(June2025)-(12.8.25).pdf

Customs General Order No. 6 of 2024: Procedure for Faceless Customs Assessment through establishment of the Central Appraising Unit (9 December 2024) [Administrative customs order]. Federal Board of Revenue, official Customs General Orders repository. https://www.fbr.gov.pk/Orders/Customs-General-Orders/131

Digital Nation Pakistan Act, 2025 (Act I of 2025; 29 January 2025). Pakistan Digital Authority, official text. https://www.pda.gov.pk/downloads/Digital-Nation-Pakistan-Act-1-of-2025.pdf

Dr Muhammad Aslam Khaki v S.S.P. (Operations), Rawalpindi, PLD 2013 SC 188.

Drug Regulatory Authority of Pakistan Act, 2012 (Act XXI of 2012), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2Fqapo%3D-sg-jjjjjjjjjjjj

Drugs Act, 1976, as amended/current at the cut-off. Drug Regulatory Authority of Pakistan, official Acts repository. https://www.dra.gov.pk/about-us/legislation/acts/

Du, M., Li, F., Zheng, G., & Srikumar, V. (2017). DeepLog: Anomaly detection and diagnosis from system logs through deep learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1285–1298). Association for Computing Machinery. https://doi.org/10.1145/3133956.3134015

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–487. https://doi.org/10.1561/0400000042

Electronic Transactions Ordinance, 2002 (Ordinance LI of 2002), as amended/current at the cut-off. Ministry of Information Technology and Telecommunication, official text. https://moitt.gov.pk/SiteImage/Legislations/ETO%20Act.pdf

Establishment of Telecommunication Appellate Tribunal Act, 2024 (Act XXV of 2024). National Assembly of Pakistan, official text. https://www.na.gov.pk/uploads/documents/6706a0c5581ba_490.pdf

Establishment of the Office of Ombudsman for the Province of Sindh Act, 1991 (Act I; published 23 January 1992), as amended/current at the cut-off. Sindh Code, official text. https://www.sindhlaws.gov.pk/LawDocument.aspx?Lang=EN&PubID=PUB-15-000768

Establishment of the Office of Wafaqi Mohtasib (Ombudsman) Order, 1983 (President's Order No. 1 of 1983), as amended/current at the cut-off. Wafaqi Mohtasib, official legal-framework page. https://www.mohtasib.gov.pk/Detail/OGM4YzViMzItYzUzMS00MDVjLWJkZjAtZTEyMjBmMWY3NWVm

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence* [Comparative foreign legislation; not Pakistani law; original Official Journal text published 12 July 2024]. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A32024R1689

Fairwork. (2023). *Fairwork Pakistan ratings 2023: Labour standards in the platform economy* [Bounded study of six platforms in Islamabad and Rawalpindi; not a nationwide or current-system census]. University of Oxford. https://fair.work/en/fw/publications/fairwork-pakistan-ratings-2023-labour-standards-in-the-platform-economy/

Faiz, A. (2023). Center-Sindh relations in Pakistan after the 18th Amendment: Anatomy of executive federalism under the PTI government, 2018–2022. Journal of Sindhi Studies, 3(2), 1–19. https://doi.org/10.1163/26670925-bja10015

Farooqi, S., & Forbes, T. (2020). Enacted discretion: Policy implementation, local government reform and education services in Pakistan. Public Management Review, 22(8), 1217–1239. https://doi.org/10.1080/14719037.2019.1630134

Federal Board of Revenue. (2024, December 14). *FBR to implement Faceless Customs Assessment System at Karachi from 15th Dec, 2024* [Official geographically bounded implementation announcement; not model-performance evidence]. https://fbr.gov.pk/fbr-to-implement-faceless-customs-assessment-system-at-karachi-from-15th-dec-2024/174183

Federal Board of Revenue. (2025a, February 1). *FBR successfully foils an attempt to game the Faceless Customs Assessment System; suspends licences of dozens of agents; arrests three persons and initiates inquiry against an appraising officer* [Official incident/enforcement announcement]. https://www.fbr.gov.pk/fbr-successfully-foils-an-attempt-to-game-the-faceless-customs-assessment-system-suspends-licences-/174198

Federal Board of Revenue. (2025b). Revenue Division Year Book 2024–25 [Official institutional report; compliance-risk management system discussed at p. 37]. https://download1.fbr.gov.pk/Docs/20251029101050494 45FBRREVENUEDIVISIONYEARBOOK2024-25.pdf

Federal Board of Revenue. (n.d.). *Web Based One Customs (WeBOC) system of goods declaration and clearance* [Official service description]. https://www.fbr.gov.pk/categ/register-customs/51149/70849/%20131192

Federal Ombudsmen Institutional Reforms Act, 2013 (Act XIV of 2013), as amended/current at the cut-off. National Assembly of Pakistan, official Gazette text. https://na.gov.pk/uploads/documents/1370389378_309.pdf

Fredrikson, M., Jha, S., & Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 1322–1333). Association for Computing Machinery. https://doi.org/10.1145/2810103.2813677

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723

General Clauses Act, 1897, as amended/current at the cut-off, especially s. 24A. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-cZ0%3D-sg-jjjjjjjjjjjj

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://www.deeplearningbook.org/

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In *3rd International Conference on Learning Representations (ICLR 2015)*. https://doi.org/10.48550/arXiv.1412.6572

Government of Pakistan, Cabinet Division. (2025). Rules of Business, 1973 (as amended up to 24 March 2025). https://www.cabinet.gov.pk/SiteImage/Misc/files/ROB%2C%201973/ROB1973-24-March-2025.pdf

Government of Pakistan, Finance Division. (2022). Performance based budget FYs 2022–23 to 2024–25. https://www.finance.gov.pk/budget/Budget_2022_23/Performance_Based_Budget_FY_2022_23_to_2024_25.pdf

Government of Pakistan, Finance Division. (2025). *Financial Management and Powers of Principal Accounting Officers Regulations, 2021* (as amended through 27 October 2025). https://www.finance.gov.pk/budget/fmp_paos_reguations_2021_amended_27102025.pdf

Government of Pakistan, Ministry of Information Technology and Telecommunication. (2022). *Pakistan Cloud First Policy* [Approved executive policy; approved 18 February 2022]. Official status register: https://moitt.gov.pk/Policies

Government of Pakistan, Ministry of Information Technology and Telecommunication. (2023, May). *Personal Data Protection Bill* [Draft bill; not enacted at the cut-off]. https://moitt.gov.pk/SiteImage/Legislations/Final%20Draft%20Personal%20Data%20Protection%20Bill%20May%202023%20(1).pdf (Official status register: https://moitt.gov.pk/Legislations)

Government of Pakistan, Ministry of Information Technology and Telecommunication. (2025). National Artificial Intelligence Policy 2025 (approved 31 July 2025). https://moitt.gov.pk/SiteImage/Policies/National%20AI%20Policy.pdf

Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face Recognition Vendor Test (FRVT), Part 3: Demographic effects* (NISTIR 8280) [Comparative technical evaluation; not evidence of any Pakistani system's performance]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8280

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning* (Proceedings of Machine Learning Research, Vol. 70, pp. 1321–1330). PMLR. https://proceedings.mlr.press/v70/guo17a.html

Gutiérrez, J. D. (2025). *Guidelines for the use of AI systems in courts and tribunals* [Comparative judicial-governance guidance; not Pakistani law]. UNESCO. https://doi.org/10.58338/LIEY8089

Gu, T., Liu, K., Dolan-Gavitt, B., & Garg, S. (2019). BadNets: Evaluating backdooring attacks on deep neural networks. *IEEE Access, 7*, 47230–47244. https://doi.org/10.1109/ACCESS.2019.2909068

Hashmi, Z. (2021). Making reliable persons: Managing descent and genealogical computation in Pakistan. *Comparative Studies in Society and History, 63*(4), 948–978. https://doi.org/10.1017/S001041752100030X

Higher Education Commission. (n.d.). *Anti-plagiarism policy* [Administrative policy; not evidence of automated merits adjudication]. https://www.hec.gov.pk/english/policies/Documents/Plagiarism-Policy.pdf

Higher Education Commission Ordinance, 2002 (Ordinance LIII of 2002; 11 September 2002), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-ap%2BZbA%3D%3D-sg-jjjjjjjjjjjj

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*(5786), 504–507. https://doi.org/10.1126/science.1127647

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hong, C. (2022). "Safe cities" in Pakistan: Knowledge infrastructures, urban planning, and the security state. *Antipode, 54*(5), 1476–1496. https://doi.org/10.1111/anti.12799

I. *A. Sherwani v Government of Pakistan*, 1991 SCMR 1041.

Income Tax Ordinance, 2001 (Ordinance XLIX of 2001), as amended up to 31 July 2025. Federal Board of Revenue, official consolidated text. https://download1.fbr.gov.pk/Docs/2025881983148210Income-Tax-Ordinance,-2001-Amended-upto-31.07.2025.pdf

Industrial Relations Act, 2012 (Act X of 2012), as amended/current within its federal scope at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2FqaZk%3D-sg-jjjjjjjjjjjj

Insurance Ordinance, 2000 (Ordinance XXXIX of 2000), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2Npa5pkaA%3D%3D-sg-jjjjjjjjjjjj

International Organization for Standardization. (2022). *ISO/IEC 27001:2022: Information security, cybersecurity and privacy protection—Information security management systems—Requirements* [Voluntary international standard; not Pakistani law unless given effect through an applicable legal, regulatory, administrative, or contractual route]. https://www.iso.org/standard/27001

International Organization for Standardization. (2023a). *ISO/IEC 23894:2023: Information technology—Artificial intelligence—Guidance on risk management* [Voluntary international standard]. https://www.iso.org/standard/77304.html

International Organization for Standardization. (2023b). *ISO/IEC 42001:2023: Information technology—Artificial intelligence—Management system* [Voluntary international standard]. https://www.iso.org/standard/42001

Investigation for Fair Trial Act, 2013 (Act I of 2013; 22 February 2013), as amended/current at the cut-off. Ministry of Law and Justice, official Gazette copy. https://drs.molaw.gov.pk/pdf/The%20Investigation%20for%20Fair%20Trial%20Act%2C%202013%20%281%29.pdf

Ishfaq Ahmed v Mushtaq Ahmed, PLD 2025 SC 582; 2025 SCP 112 (Supreme Court of Pakistan, Civil Petition No. 1010-L of 2022, decided 13 March 2025). https://www.supremecourt.gov.pk/downloads_judgements/c.p._1010_l_2022.pdf

Ishtiaq Ahmed Mirza v Federation of Pakistan, PLD 2019 SC 675.

Jacobs, A. Z., & Wallach, H. (2021). Measurement and fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 375–385). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445901

Jain, S., & Wallace, B. C. (2019). Attention is not explanation. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 3543–3556). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1357

Jan, S. U., Khan, M. S. A., & Khan, A. S. (2024). Organizational readiness to adopt artificial intelligence in the library and information sector of Pakistan. *Evidence Based Library and Information Practice, 19*(1), 58–76. https://doi.org/10.18438/eblip30408

Kazi, A. M., Qazi, S. A., Ahsan, N., Khawaja, S., Sameen, F., Saqib, M., Khan Mughal, M. A., Wajidali, Z., Ali, S., Ahmed, R. M., Kalimuddin, H., Rauf, Y., Mahmood, F., Zafar, S., Abbasi, T. A., Khoumbati, K.-U.-R., Abbasi, M. A., & Stergioulas, L. K. (2020). Current challenges of digital health interventions in Pakistan: Mixed methods analysis. Journal of Medical Internet Research, 22(9), Article e21691. https://doi.org/10.2196/21691

Khokhar, F. (2024). Privity of contract and rights of third party: An appraisal of Pakistani and English courts' judgments on the doctrine. *Policy Perspectives, 21*(1), 127–150. https://doi.org/10.13169/polipers.21.1.ra7

Khyber Pakhtunkhwa Public Procurement Regulatory Authority Act, 2012 (Khyber Pakhtunkhwa Act XI of 2012), as amended/current at the cut-off. Khyber Pakhtunkhwa Public Procurement Regulatory Authority, official legislation repository. https://www.kppra.gov.pk/kppra/download_new

Königstorfer, F., & Thalmann, S. (2022). AI documentation: A path to accountability. *Journal of Responsible Technology, 11*, 100043. https://doi.org/10.1016/j.jrt.2022.100043

Komatsu, T. (2009). Qualitative inquiry into local education administration in Pakistan. International Journal of Educational Development, 29(3), 219–226. https://doi.org/10.1016/j.ijedudev.2008.04.004

Laila Qayyum v Fawad Qayum, PLD 2019 SC 449. https://gdpakistan.org/wp-content/uploads/2020/10/DNA-test-denied_Redacted.pdf

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Legal Aid and Justice Authority Act, 2020 (Act XVI of 2020). Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2NpaZxr-sg-jjjjjjjjjjjj

Legal Practitioners and Bar Councils Act, 1973 (Act XXXV of 1973), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2Npa5prbw%3D%3D-sg-jjjjjjjjjjjj

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*. Curran Associates. https://proceedings.neurips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

Majoka, Z., Wieser, C., Qazi, M., Guzman Fonseca, D., Sohnesen, T. P., & Khan, I. (2024). *Mind the gap: Assessing Pakistan's National Socio-Economic Registry (NSER)*. World Bank. https://doi.org/10.1596/42566

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (Proceedings of Machine Learning Research, Vol. 54, pp. 1273–1282). PMLR. https://proceedings.mlr.press/v54/mcmahan17a.html

Meera Shafi v Ali Zafar, PLD 2023 SC 211 (Supreme Court of Pakistan, Civil Petition No. 1795 of 2022, decided 21 November 2022). https://www.supremecourt.gov.pk/downloads_judgements/c.p.1795_2022.pdf

Messrs Mastersons through its Partner v Messrs Ebrahim Enterprises and another, 1988 CLC 1381 (Sindh High Court, Pakistan).

Mian Khalid Pervaiz v State, 2021 SCMR 522.

Ministry of Information Technology and Telecommunication. (2021). *National Cyber Security Policy 2021* [Approved executive policy; official register records approval on 27 July 2021]. https://pkcert.gov.pk/uploads/2023/06/2021-National-Cyber-Security-Policy-Final.pdf (Official status register: https://moitt.gov.pk/Policies)

Ministry of Information Technology and Telecommunication. (n.d.-a). *Legislations* [Official status register used to verify the draft status of the Personal Data Protection Bill at the cut-off]. https://moitt.gov.pk/Legislations

Ministry of Information Technology and Telecommunication. (n.d.-b). *Policies* [Official status register used to verify policy approval and draft status]. https://moitt.gov.pk/Policies

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287596

Mohtarma Benazir Bhutto v President of Pakistan, PLD 1998 SC 388.

Mustafa Impex, Karachi v Government of Pakistan, PLD 2016 SC 808.

National Archives Act, 1993 (Act VI of 1993), original enacted text. National Assembly of Pakistan, official text. https://www.na.gov.pk/uploads/documents/1329724034_276.pdf

National Assembly of Pakistan. (n.d.). *Acts of Parliament* [Official enactment register consulted for the statutory position at the cut-off]. https://www.na.gov.pk/en/acts.php

National Assembly of Pakistan, Public Accounts Committee. (n.d.). *Rules relating to the Public Accounts Committee* (especially r. 203). https://www.pac.na.gov.pk/?q=rules-relating-pac

National Commission for Human Rights Act, 2012 (Act XVI of 2012; 5 June 2012), original enacted text. Official Gazette copy hosted by the Sindh Human Rights Commission. https://www.shrc.org.pk/downloads/laws/Pakistan/15-National-Commission-for-Human-Rights-Act-2012.pdf

National Cyber Security Centre. (2023, November 27). *Guidelines for secure AI system development* (Version 1.0) [Joint international guidance; not Pakistani law]. https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development

National Database and Registration Authority (Digital Identity) Regulations, 2025 (S.R.O. 1956(I)/2025; notified 13 October and published/effective 14 October 2025). Ministry of Law and Justice, official Gazette copy. https://drs.molaw.gov.pk/pdf/Digital_Identity_Regulations_S.%20R.%20O._1956_2025_1.pdf

National Database and Registration Authority (National Data Exchange Layer) Regulations, 2025 (S.R.O. 1957(I)/2025; gazetted 14 October 2025). Ministry of Law and Justice, official Gazette copy. https://drs.molaw.gov.pk/pdf/NDEL_Regulations_S.%20R.%20O._1957_2025..pdf

National Database and Registration Authority (National Data Repository) Regulations, 2024 (S.R.O. 257(I)/2025; gazetted 6 March 2025). Ministry of Law and Justice, official Gazette copy. https://drs.molaw.gov.pk/pdf/NADRA%20(National%20Data%20Repository)%20Regulations%202024.pdf

National Database and Registration Authority. (n.d.-a). *Customer support and complaints*. https://www.nadra.gov.pk/customer-support

National Database and Registration Authority. (n.d.-b). *Inclusive registration*. https://www.nadra.gov.pk/inclusiveRegistration

National Database and Registration Authority. (n.d.-c). *National identity card modification services and fee structure*. https://www.nadra.gov.pk/feeStructure

National Database and Registration Authority Ordinance, 2000 (Ordinance VIII of 2000), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/pdffiles/administrator413b54beecc75f8df45c9a57fcea6d16.pdf

National Information Technology Board Act, 2022 (Act XVII of 2022). Official Gazette copy hosted by the Senate of Pakistan. https://senate.gov.pk/uploads/documents/1661158094_763.pdf

National Logistic Cell v Irfan Khan, 2015 SCMR 1406 (Supreme Court of Pakistan).

National Security Agency, Cybersecurity and Infrastructure Security Agency, Federal Bureau of Investigation, Australian Signals Directorate's Australian Cyber Security Centre, New Zealand Government Communications Security Bureau's National Cyber Security Centre, & United Kingdom National Cyber Security Centre. (2025, May). *AI data security: Best practices for securing data used to train and operate AI systems* (Version 1.0) [Joint international cybersecurity guidance; not Pakistani law]. https://media.defense.gov/2025/May/22/2003720601/-1/-1/0/CSI_AI_DATA_SECURITY.PDF

Near, J. P., Darais, D., Lefkovitz, N., & Howarth, G. S. (2025). Guidelines for evaluating differential privacy guarantees (NIST Special Publication 800-226) [Comparative technical guidance; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-226

Nelson, A., Rekhi, S., Souppaya, M., & Scarfone, K. (2025). *Incident response recommendations and considerations for cybersecurity risk management: A CSF 2.0 community profile* (NIST Special Publication 800-61, Rev. 3) [Comparative guidance; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-61r3

Organisation for Economic Co-operation and Development. (2022). *OECD framework for the classification of AI systems* (OECD Digital Economy Papers No. 323) [Comparative classification framework; not Pakistani law]. https://doi.org/10.1787/cb6d9eca-en

Organisation for Economic Co-operation and Development. (2024a, March 5). Explanatory memorandum on the updated OECD definition of an AI system (OECD Artificial Intelligence Papers No. 8). https://doi.org/10.1787/623da898-en

Organisation for Economic Co-operation and Development. (2024b). *Recommendation of the Council on artificial intelligence* (OECD/LEGAL/0449; adopted 2019, amended 2024) [International soft-law recommendation; not Pakistani law]. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Organisation for Economic Co-operation and Development. (2025). *AI in public procurement*. In *Governing with artificial intelligence* [Comparative policy guidance; not Pakistani procurement law]. https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en/full-report/ai-in-public-procurement_2e095543.html

Pakistan Bar Council. (1976). *Pakistan Legal Practitioners and Bar Councils Rules, 1976* (ch. XII, Canons of Professional Conduct and Etiquette of Advocates), as amended/current at the cut-off. https://pakistanbarcouncil.org/storage/acts/Bar%20Council%20Rules.pdf

Pakistan Bureau of Statistics. (2021, May). *Pakistan Social and Living Standards Measurement Survey 2019–20: Provincial / district*. https://www.pbs.gov.pk/wp-content/uploads/2020/07/PSLM_2019_20_District_Level.pdf

Pakistan Citizenship Act, 1951 (Act II of 1951), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-ap%2BXZQ%3D%3D-sg-jjjjjjjjjjjj

Pakistan Customs, Directorate General of Intelligence & Risk Management, Risk Management Unit. (2025). *Pakistan Customs risk management system* [Institutional presentation; reports IREE implementation and an RMS 2.0 simulation, not independently verified production operation]. WTO Trade Facilitation Agreement Database. https://tfadatabase.org/en/uploads/thematicdiscussiondocument/pakistan_presentation_on_rms_at_wto_final.pdf

Pakistan Medical and Dental Council Act, 2022 (Act IV of 2023; 12 January 2023), as amended/current at the cut-off. Pakistan Medical and Dental Council, official text. https://pmdc.pk/Documents/law/PMDC%20Act.pdf

Pakistan Single Window. (2025a). PSW launches "Khizer" – Pakistan's first AI-powered trade chatbot. The View: PSW Quarterly Newsletter, July–September 2025 [Institutional launch report; not independent performance validation]. https://www.psw.gov.pk/view-psw-quarterly-newsletter-Sept-2025-Khizer

Pakistan Single Window. (2025b). The View: Monthly update, October 2025 [Institutional report identifying Khizer integration and Sybrid Private Limited collaboration]. https://www.psw.gov.pk/media/newsletter/the-view-october-2025.pdf

Pakistan Telecommunication Authority. (2021). *Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards) Rules, 2021* [Delegated legislation within its statutory scope]. https://www.pta.gov.pk/assets/media/removal_blocking_unlawful_content_rules_2021_20102021.pdf

Pakistan Telecommunication Authority. (2023, February 14). *PTA launches nTSOC to strengthen cybersecurity in Pakistan* [Official announcement of a platform and stated integrations; not evidence of a neural-network component or model performance]. https://www.pta.gov.pk/category/pta-launches-ntsoc-to-strengthen-cybersecurity-in-pakistan-967073847-2023-06-01

Pakistan Telecommunication Authority. (2025). *Critical Telecom Data and Infrastructure Security Regulations 2025 (CTDISR-2025)* (notification dated 29 October 2025) [Delegated telecom regulations; commenced on Gazette notification; the official register lists them as current and the 2020 regulations as repealed]. https://www.pta.gov.pk/assets/media/2025-10-30-CTDISR-2025.pdf (Official status register: https://www.pta.gov.pk/category/regulations-1674158059-2023-05-30)

Pakistan Telecommunication (Re-organization) Act, 1996 (Act XVII of 1996), consolidated with amendments including Act XXV of 2024. Pakistan Telecommunication Authority, official legislation page. https://www.pta.gov.pk/category/act-1397047816-2023-05-30 (Pakistan Code record: https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apqWaw%3D%3D-sg-jjjjjjjjjjjj)

Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys, 54*(2), Article 38, 1–38. https://doi.org/10.1145/3439950

Parlov, N., Mateša, B., & Mladinić, A. (2025). Structuring AI risk management framework: EU AI Act FRIA, GDPR DPIA and ISO 42001/23894. In 2025 14th Mediterranean Conference on Embedded Computing (MECO) (pp. 1–8). IEEE. https://doi.org/10.1109/MECO66322.2025.11049196

Pascoe, C., Quinn, S., & Scarfone, K. (2024). *The NIST Cybersecurity Framework (CSF) 2.0* (NIST CSWP 29) [Voluntary comparative framework; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.CSWP.29

Pawlicki, M., Kozik, R., & Choraś, M. (2022). A survey on neural networks for (cyber-)security and (cyber-)security of neural networks. *Neurocomputing, 500*, 1075–1087. https://doi.org/10.1016/j.neucom.2022.06.002

Payment Systems and Electronic Fund Transfers Act, 2007, as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2FsaZY%3D-sg-jjjjjjjjjjj

Pervaiz, Z., Akram, S., & Ahmad Jan, S. (2021). Social exclusion in Pakistan: An ethnographic and regional perspective. *International Journal of Sociology and Social Policy, 41*(11/12), 1183–1194. https://doi.org/10.1108/IJSSP-01-2021-0001

Pfister, J. (2022, February 24). Automation bias & the proposal of the EU-regulation on AI. Jusletter IT. https://doi.org/10.38023/6ec715d9-0ed6-4a44-8be0-6874a866d8e8

Phillips, P. J., Hahn, C. A., Fontana, P. C., Yates, A. N., Greene, K. K., Broniatowski, D. A., & Przybocki, M. A. (2021). *Four principles of explainable artificial intelligence* (NISTIR 8312) [Comparative technical guidance; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8312

Prevention of Electronic Crimes Act, 2016 (Act XL of 2016), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/pdffiles/administrator6a061efe0ed5bd153fa8b79b8eb4cba7.pdf

Prevention of Electronic Crimes (Amendment) Act, 2025 (Act II of 2025). Senate of Pakistan, official text. https://senate.gov.pk/uploads/documents/1738226500_897.pdf

Public Finance Management Act, 2019, as amended through 30 June 2024. Government of Pakistan, Finance Division, official consolidated text. https://finance.gov.pk/budget/pfm_act_2019_30june2024.pdf

Public Procurement Regulatory Authority. (2024a). *National Standard Bidding Document for Procurement of Cloud Services* (S.R.O. 730(I)/2024). https://ppra.gov.pk/sbd

Public Procurement Regulatory Authority. (2024b). *Regulations on "Mechanism for Blacklisting and Debarment of Bidders or Contractors Regulations, 2024"* (S.R.O. 460(I)/2024; notification dated 28 March 2024; Gazette publication 1 April 2024). https://ppra.gov.pk/regulations

Public Procurement Regulatory Authority (Amendment) Act, 2023 (Act II of 2023; 13 January 2023). National Assembly of Pakistan, official text. https://www.na.gov.pk/uploads/documents/63da17a108f2b_452.pdf

Public Procurement Regulatory Authority. (n.d.-a). *Blacklisted/debarred firms*. https://epms.ppra.gov.pk/public/active-blacklisted-firms

Public Procurement Regulatory Authority. (n.d.-b). *National standard bidding documents*. https://old.ppra.org.pk/sbd.asp

Public Procurement Regulatory Authority Ordinance, 2002 (Ordinance XXII of 2002; 15 May 2002), as amended. Pakistan Code, official text. https://pakistancode.gov.pk/pdffiles/administrator28cd90a18aea57042f8ddc40ccd8cf98.pdf (Official PPRA consolidated page: https://ppra.gov.pk/ordinance)

Public Procurement Rules, 2004 (S.R.O. 432(I)/2004; 8 June 2004), as amended through S.R.O. 763(I)/2025 and current at the cut-off. Public Procurement Regulatory Authority, official consolidated text. https://epms.ppra.gov.pk/public/procurement-rules

Punjab Consumer Protection Act, 2005, as amended/current at the cut-off. Directorate of Consumer Protection Council, Government of Punjab, official statutory-powers page. https://pcpc.punjab.gov.pk/power_authority

Punjab Healthcare Commission Act, 2010 (Act XVI of 2010), as amended/current at the cut-off. Punjab Laws, official register. https://punjablaws.punjab.gov.pk/en/articles_by_category/53

Punjab Information Technology Board. (2025, May 20). PITB conducts a field visit to Faisalabad Waste Management Company [Official report of facial-attendance use at observed Faisalabad sites]. https://pitb.gov.pk/node/10743

Punjab Information Technology Board. (n.d.). *Online College Admissions System* [Operational digital application channel; not evidence of AI merits adjudication]. https://ocas.punjab.gov.pk/ocas

Punjab Land Records Authority Act, 2017 (Punjab Act VI of 2017). Punjab Laws, official Gazette copy. https://www.punjablaws.gov.pk/GazetteNotifications/222537.pdf

Punjab Land Records Authority. (n.d.-a). *Correction of land records* [Official service route]. https://www.punjab-zameen.gov.pk/correctionRecordsInfo

Punjab Land Records Authority. (n.d.-b). *Punjab Land Records Authority* [Official service portal; evidence of digital land-record infrastructure, not AI title or valuation]. https://www.punjab-zameen.gov.pk/

Punjab Office of the Ombudsman Act, 1997 (Act X of 1997), as amended/current at the cut-off. Punjab Laws, official text. https://www.punjablaws.gov.pk/title.php

Punjab Police. (2023, July 26). *Punjab Police launch AI-powered system to track criminals* [Official provincial launch claim; not model-performance or evidential-reliability proof]. https://punjabpolice.gov.pk/node/13893

Punjab Police. (2024a, February 23). *PITB's AI-powered system for Punjab police set to curb crimes [The Nation report reproduced on the official Punjab Police website].* https://punjabpolice.gov.pk/node/17067

Punjab Police. (2024b, September 29). *Punjab Safe Cities: Enhancing security and efficiency* [Official programme description]. https://punjabpolice.gov.pk/node/20831

Punjab Procurement Rules, 2014, as amended up to 30 April 2025. Punjab Public Procurement Regulatory Authority, official consolidated text. https://ppra.punjab.gov.pk/system/files/Punjab%20Procurement%20Rules%202014%20%28amended%20upto%2030-04-2025%29.pdf

Punjab Safe Cities Authority Act, 2016 (Punjab Act I of 2016). Punjab Laws, official text. https://www.punjablaws.gov.pk/lawspdf/2619.pdf

Qanun-e-Shahadat Order, 1984 (President's Order No. 10 of 1984), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2Npa5plaw%3D%3D-sg-jjjjjjjjjjjj

Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (Eds.). (2009). *Dataset shift in machine learning*. MIT Press. https://mitpress.mit.edu/9780262170055/dataset-shift-in-machine-learning/

Raff, E., Barker, J., Sylvester, J., Brandon, R., Catanzaro, B., & Nicholas, C. (2018). Malware detection by eating a whole EXE. In *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence* (pp. 268–276). Association for the Advancement of Artificial Intelligence. https://aaai.org/papers/268-ws0432-aaaiw-18-16422/

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372873

Rasool, N., & Rasool, M. (2022). Challenges for expert evidence in the justice system of Pakistan. Journal of Forensic Science and Medicine, 8(2), 62–67. https://doi.org/10.4103/jfsm.jfsm_16_21

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2016 O.J. (L 119) 1 [Comparative foreign legislation; territorial and material application must be established on its own terms]. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778

Riccio, G. M. (2025). Tort rules and regulatory sandboxes: The influence of the national jurisdictions. In 2025 MIPRO 48th ICT and Electronics Convention (pp. 1598–1603). IEEE. https://doi.org/10.1109/MIPRO65660.2025.11131991

Right of Access to Information Act, 2017, as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2Noa5c%3D-sg-jjjjjjjjjjjj

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020, August). *Zero trust architecture* (NIST Special Publication 800-207) [Comparative guidance; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-207

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536. https://doi.org/10.1038/323533a0

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Salahuddin and 2 others v Frontier Sugar Mills and Distillery Ltd., Takht Bhai and 10 others, PLD 1975 SC 244 (Supreme Court of Pakistan).

Scarfone, K., & Mell, P. (2007, February). *Guide to intrusion detection and prevention systems (IDPS)* (NIST Special Publication 800-94) [Comparative guidance; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-94

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence* (NIST Special Publication 1270) [Comparative technical guidance; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1270

Schwarzschild, A., Goldblum, M., Gupta, A., Dickerson, J. P., & Goldstein, T. (2021). Just how toxic is data poisoning? A unified benchmark for backdoor and data poisoning attacks. In *Proceedings of the 38th International Conference on Machine Learning* (Proceedings of Machine Learning Research, Vol. 139, pp. 9389–9398). PMLR. https://proceedings.mlr.press/v139/schwarzschild21a.html

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 28* (pp. 2503–2511). Curran Associates. https://papers.nips.cc/paper_files/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

Securities Act, 2015, as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2JwZp8%3D-sg-jjjjjjjjjjj

Securities and Exchange Commission of Pakistan. (2019). *SECP Regulatory Sandbox Guidelines, 2019* [Sectoral sandbox guidelines; not a general exemption from applicable law]. https://www.secp.gov.pk/document/secp-regulatory-sandbox-guidelines-2019/

Securities and Exchange Commission of Pakistan. (2022). *Circular No. 15 of 2022: Requirements for NBFCs engaged in digital lending* [Sectoral regulatory circular]. https://www.secp.gov.pk/document/circular-no-15-of-2022-requirements-for-nbfcs-engaged-in-digital-lending/

Securities and Exchange Commission of Pakistan. (2023, February 28). *SECP organises financial reporting workshop on NBMF sector* [Press release containing 2022 digital-lending disbursement information; not model-level evidence]. https://www.secp.gov.pk/wp-content/uploads/2023/03/Press-Release-February-28-SECP-organizes-financial-reporting-workshop-on-NBMF-sector.pdf

Securities and Exchange Commission of Pakistan. (2024, May 16). *SECP enables digital lenders to launch novel solutions* [Press release concerning Circular No. 12 of 2024; not evidence of every lender's model or outcome]. https://www.secp.gov.pk/wp-content/uploads/2024/05/SECP-enables-digital-lenders-to-launch-novel-solutions.pdf

Securities and Exchange Commission of Pakistan Act, 1997 (Act XLII of 1997), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw2-apaUY2Fqa-apqbYw%3D%3D-sg-jjjjjjjjjjjj-con-15784

Securities and Exchange Commission of Pakistan. (n.d.). *Regulatory sandbox* [Official programme information]. https://www.secp.gov.pk/regulatory-sandbox/

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287598

Service Tribunals Act, 1973 (Act LXX of 1973), as amended/current at the cut-off. Pakistan Code, official text. https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-bpuUY2px-sg-

Shahid, S. Z., Bangash, M. A., Hussain, M. R., & Arshad, S. (2024). CyberCure Malware defense system. In *2024 4th International Conference on Innovations in Computer Science (ICONICS)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICONICS64289.2024.10824664

Shehla Zia v WAPDA, PLD 1994 SC 693.

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy* (pp. 3–18). IEEE. https://doi.org/10.1109/SP.2017.41

Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence, 2*(1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792

Sindh Public Procurement Rules, 2010, as amended/current at the cut-off. Sindh Public Procurement Regulatory Authority, official text. https://www.pprasindh.gov.pk/downloads/files/FINALSPPRARULES08022010.pdf

Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 180–186). Association for Computing Machinery. https://doi.org/10.1145/3375627.3375830

Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE Symposium on Security and Privacy* (pp. 305–316). IEEE. https://doi.org/10.1109/SP.2010.25

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958. https://www.jmlr.org/papers/v15/srivastava14a.html

State Bank of Pakistan. (2017, May 30). *Enterprise Technology Governance and Risk Management Framework for Financial Institutions* (BPRD Circular No. 05 of 2017) [Sectoral regulatory framework]. https://www.sbp.org.pk/circulars/bprd-circular-no-05-of-2017

State Bank of Pakistan. (2023, April 14). *Measures to Enhance Security of Digital Banking Products and Services* (BPRD Circular No. 04 of 2023) [Sectoral regulatory circular]. https://www.sbp.org.pk/circulars/bprd-circular-no-04-of-2023

State Bank of Pakistan. (2024, December 17). *Launch of new eCIB System (V2)* (CPD Circular Letter No. 02 of 2024) [Sectoral circular]. https://www.sbp.org.pk/CPD/2024/CL2.pdf

State Bank of Pakistan. (2025a). *Guidelines for Regulatory Sandbox* [Sectoral sandbox guidance; not a general exemption from applicable law]. https://www.sbp.org.pk/our-operations/innovation/regulatory-sandbox

State Bank of Pakistan. (2025b, October 3). *Technology Risk Management Framework for Payment Institutions* (PSP&OD Circular No. 04 of 2025) [Sectoral regulatory framework applicable within the circular's stated scope]. https://www.sbp.org.pk/circulars/pspod-circular-no-04-of-2025

State Bank of Pakistan Act, 1956 (Act XXXIII of 1956), as amended/current at the cut-off. Ministry of Law and Justice, official document history. https://drs.molaw.gov.pk/documents/document-versions?docId=1688

State Bank of Pakistan. (n.d.). *Frequently asked questions: Electronic Credit Information Bureau (eCIB)* [Official explanation of eCIB and the lender's retained credit-decision role]. https://www.sbp.org.pk/faqs/faqs-credit-information-bureau-cib

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning* (Proceedings of Machine Learning Research, Vol. 70, pp. 3319–3328). PMLR. https://proceedings.mlr.press/v70/sundararajan17a.html

Suresh, H., & Guttag, J. V. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Article 17, pp. 1–9). Association for Computing Machinery. https://doi.org/10.1145/3465416.3483305

Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1) [Voluntary comparative framework; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Tahir, M. H., Shams, S., Fiaz, L., Adeeba, F., & Hussain, S. (2025). Benchmarking the performance of pre-trained LLMs across Urdu NLP tasks. In K. Sarveswaran, A. Vaidya, B. K. Bal, S. Shams, & S. Thapa (Eds.), *Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025)* (pp. 17–34). International Committee on Computational Linguistics. https://aclanthology.org/2025.chipsal-1.3/

Talaat Inayatullah Khan and another v Dr Anis Ahmad Sheikh, PLD 2015 Sindh 134 (Sindh High Court, Pakistan).

Tareen, A., & Tareen, K. I. (2016). Mental health law in Pakistan. BJPsych International, 13(3), 67–69. https://doi.org/10.1192/S2056474000001276

Tartaro, A. (2023). Regulating by standards: Current progress and main challenges in the standardisation of artificial intelligence in support of the AI Act. European Journal of Privacy Law & Technologies, (1), 147–174. https://doi.org/10.57230/EJPLT222AT

Temoshok, D., Fenton, J. L., Choong, Y.-Y., Lefkovitz, N., Regenscheid, A., Galluzzo, R., & Richer, J. P. (2025, July). *Digital identity guidelines: Authentication and authenticator management* (NIST Special Publication 800-63B-4) [Comparative guidance; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-63B-4

Tramèr, F., Zhang, F., Juels, A., Reiter, M. K., & Ristenpart, T. (2016). Stealing machine learning models via prediction APIs. In *25th USENIX Security Symposium (USENIX Security 16)* (pp. 601–618). USENIX Association. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/tramer

Umer, L., Khan, M. H., & Ayaz, Y. (2023). Transforming healthcare with artificial intelligence in Pakistan: A comprehensive overview. Pakistan Armed Forces Medical Journal, 73(4), 955–963. https://doi.org/10.51253/pafmj.v73i4.10852

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence* (SHS/BIO/REC-AIETHICS/2021) [International recommendation; not Pakistani law]. https://unesdoc.unesco.org/ark:/48223/pf0000380455

UNESCO. (2023). *Ethical impact assessment: A tool of the Recommendation on the Ethics of Artificial Intelligence* [Comparative assessment methodology; not Pakistani law]. https://unesdoc.unesco.org/ark:/48223/pf0000386276

United Nations High Commissioner for Refugees. (2022, June 3). *Government, UNHCR verify 1.3 million Afghan refugees in Pakistan*. https://www.unhcr.org/pk/news/stories/government-unhcr-verify-1-3-million-afghan-refugees-pakistan

United States Government Accountability Office. (2021). *Artificial intelligence: An accountability framework for federal agencies and other entities* (GAO-21-519SP) [Comparative audit framework; not Pakistani law]. https://www.gao.gov/products/gao-21-519sp

Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). *Adversarial machine learning: A taxonomy and terminology of attacks and mitigations* (NIST AI 100-2e2025) [Comparative technical report; not Pakistani law]. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-2e2025

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 5998–6008). Curran Associates. https://papers.nips.cc/paper/7181-attention-is-all-you-need

Virtual Assets Ordinance, 2025 (Ordinance VII of 2025; Gazette publication 9 July 2025). National Assembly of Pakistan, official Gazette text. https://na.gov.pk/uploads/documents/68babbb0b32b8_903.pdf

Vázquez Matilla, F. J. (2025). La inteligencia artificial en la contratación pública: Marco jurídico para su adquisición y aplicación en los procesos de compra [Artificial intelligence in public procurement: Legal framework for its acquisition and application in purchasing processes]. *IUS ET VERITAS*, (71), 1–16. https://doi.org/10.18800/iusetveritas.202502.008

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*(2), 841–887. https://jolt.law.harvard.edu/assets/articlePDFs/v31/Counterfactual-Explanations-without-Opening-the-Black-Box-Sandra-Wachter-et-al.pdf

Warid Telecom (Pvt) Ltd v Pakistan Telecommunication Authority, 2015 SCMR 338.

Waseem, M. (2006). Constitutionalism in Pakistan: The changing patterns of dyarchy. Diogenes, 53(4), 102–115. https://doi.org/10.1177/0392192106070361

Wexler, R. (2018). Life, liberty, and trade secrets: Intellectual property in the criminal justice system. *Stanford Law Review, 70*(5), 1343–1429. https://review.law.stanford.edu/wp-content/uploads/sites/3/2018/06/70-Stan.-L.-Rev.-1343.pdf

World Bank. (2017). *Pakistan—Punjab Land Records Management and Information Systems Project: Implementation completion and results report.* https://documents1.worldbank.org/curated/en/632241498842804246/pdf/ICR00003719.pdf

World Bank. (2021). *GovTech procurement practice note* [Comparative procurement guidance; not Pakistani procurement law]. https://documents1.worldbank.org/curated/en/832711611728574814/pdf/GovTech-Procurement-Practice-Note.pdf

World Bank. (2021). *Stakeholder engagement plan: Pakistan Crisis-Resilient Social Protection (CRISP) Project* [Programme grievance and eligibility-review material]. https://documents1.worldbank.org/curated/en/983971611836024303/pdf/Stakeholder-Engagement-Plan-SEP-Pakistan-Crisis-Resilient-Social-Protection-CRISP-P174484.pdf

World Bank. (2023). *Punjab Urban Land Systems Enhancement Project: Project paper*. https://documents1.worldbank.org/curated/en/099090723115518579/pdf/P1729450018b96030b4930665a79d9d3c1.pdf

World Bank. (2024). *Crisis-Resilient Social Protection—Additional Financing (P181558): Program paper* (Report No. PAD5619). https://documents1.worldbank.org/curated/en/099053124122038063/pdf/BOSIB1c8dae7ae03d181981243d144d13c8.pdf

Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. In *Advances in Neural Information Processing Systems 32*. Curran Associates. https://papers.nips.cc/paper_files/paper/2019/hash/60a6c4002cc7b29142def8871531281a-Abstract.html