IAIGCRM

INSTITUTE FOR AI GOVERNANCE, CYBERSECURITY & RISK MANAGEMENT

RESEARCH ARTICLE / RESEARCH VERSION 0.3

Protecting Personal Information in AI Systems Across Geographic and Regulatory Boundaries

Governance, Cybersecurity and Risk Management for AI Data

Research version v0.3
Published 29 October 2026

Published research version. See limitations and disclosure.

Scope: personal information definitions, regional regulation, AI risks, security and privacy controls, assessments and training.

Abstract

Artificial intelligence changes how personal information is collected, linked, inferred, disclosed and retained. Traditional security controls remain necessary, but they cannot by themselves establish whether processing is legitimate, proportionate or fair. This article proposes an integrated approach connecting personal-information definitions, regional legal duties, AI-specific risks, security and privacy controls, and evidence of effective operation. It uses NIST SP 800-53 Revision 5 and ISO/IEC 27002:2022 as complementary control references, with particular attention to North America, Europe, Japan and selected South American jurisdictions. The proposed assessment sequence combines data discovery, privacy impact assessment, business impact analysis, threat modeling, adversarial testing and ongoing monitoring. The central recommendation is to govern personal information throughout the AI lifecycle, including prompts, retrieval stores, embeddings, model artifacts, logs and downstream actions.

Keywords: personal information; PII; AI governance; privacy engineering; GDPR; APPI; LGPD; NIST SP 800-53; ISO/IEC 27002.

1. Introduction: data protection as an AI governance obligation

AI systems transform personal information into classifications, recommendations, predictions and generated content.

A hospital's AI assistant grounds a discharge summary in patient records, medication lists and clinical guidance. An insurer combines claim photographs, adjuster notes and customer histories to identify possible fraud. A multinational manufacturer deploys an employee assistant that retrieves engineering documents, travel records and occupational-health policies across offices in the United States, Europe and Japan. Each use case offers practical value while creating different pathways for personal information to enter an AI system, influence an output and cross a geographic or regulatory boundary.

The same architecture can support helpful assistance and inappropriate disclosure. A hospital assistant might retrieve another patient's chart; a claims model might infer a medical condition that was never explicitly supplied; an employee agent might transmit a confidential attachment to an external tool. These outcomes need not begin with a conventional cyberattack. Excessive permissions, incompatible reuse, flawed grounding, inaccurate inferences and unclear vendor roles can produce harm during otherwise normal operation. Governance establishes responsibility and limits. Cybersecurity protects information and processing against unauthorized activity. Privacy management evaluates whether information should be processed at all, for which purposes, under whose authority and with what consequences for individuals. Risk management connects those decisions to evidence, treatment priorities and accountable acceptance of residual risk.

The article's working proposition is that effective AI data protection requires three linked capabilities: justified processing, technically enforced boundaries, and demonstrated effectiveness. An encrypted dataset can still be collected unlawfully. A properly authenticated employee can still misuse information. An accurate AI system can still create disproportionate surveillance. Conversely, a well-written privacy notice provides little protection when retrieval permissions expose every customer record.

This is a comparative practice article, not an empirical study or an exhaustive legal survey. Legal summaries are a dated research snapshot. Applicability depends on processing activities, organizational roles, affected people, sector, location and cross-border reach. The control mappings below are the Institute's proposed implementation crosswalks, not official equivalence determinations or proof of compliance.

2. What is personally identifiable information?

2.1 Working definition

For this article, **personal information means information relating to a person who is identified or reasonably identifiable, directly or through linkage, inference or combination with other available information**. PII is used as a practical umbrella term; legal definitions must still be applied separately. Some laws additionally protect household information, prescribe specific identifiers or distinguish sensitive categories.

Table 1

Source	Definition or faithful paraphrase	Implication for AI systems
NIST SP 800-122 and NIST glossary	Includes information that distinguishes or traces identity and other information linked or linkable to an individual, including health, financial, education and employment records.	Removing a name does not remove linkage risk. Evaluate combinations and context.
ISO/IEC 29100 privacy framework	Paraphrase: information that identifies a natural person, or is or may be linked to a natural person directly or indirectly. The 2024 edition supplies common privacy terminology and principles.	Include indirect identifiers and maintain an explicit identification/linkage assessment.
OECD Privacy Guidelines	Information relating to an identified or identifiable individual.	A useful international conceptual foundation; the Guidelines are not a substitute for national law.
GDPR Article 4(1)	Information relating to an identified or identifiable natural person; identification may be direct or indirect.	Online identifiers, location and identity-related factors can qualify.
Canada PIPEDA	Information about an identifiable individual, subject to the Act's scope and exclusions.	Opinions, assessments and inferences can be personal information, not just factual records.
Japan APPI	Information about a living individual identifiable from descriptions, linkage with other information, or prescribed individual identification codes.	Distinguish personal information, personal data and retained personal data; the categories carry different duties.

Sources: NIST definitions and SP 800-122; ISO/IEC 29100:2024; OECD Privacy Guidelines; GDPR; PIPEDA; APPI. See references. ISO wording is paraphrased, not reproduced as an authenticated quotation; exact clause wording requires review of the licensed edition.

2.2 Important distinctions

Direct identifiers point readily to a person: names, national identifiers and account numbers. **Quasi-identifiers** can identify through combination: birth date, postal code, employer and unusual events. **Sensitive information** is partly a legal category and partly an operational risk classification; those two meanings must not be conflated. **Inferred information** includes predicted health status, purchasing propensity or vulnerability. An inference can be personal information even when uncertain or wrong.

Pseudonymization replaces identifiers while preserving a route to linkage. Under GDPR, pseudonymized information remains personal data when attribution remains possible. **Anonymization** requires a contextual assessment of identification risk, not simply deleting a name. Embeddings, synthetic records and model weights must be assessed for possible linkage, memorization or disclosure; they are not automatically anonymous, and not every model artifact necessarily constitutes personal data.

Public availability does not create unrestricted permission to train a model. Likewise, collecting data for service delivery does not automatically authorize fine-tuning, advertising or unrelated profiling.

2.3 Personal information across the AI data lifecycle

AI data is broader than the dataset used to train a model. Governance must follow both source information and the representations derived from it. An operational inventory should distinguish the following layers.

Table 2

AI data layer	Meaning and examples	Data protection implications
ML training data	Records used to fit model parameters, including supervised examples, labels, images and text	Verify source authority, collection purpose and suitability for training; minimize personal content and evaluate memorization.
Fine-tuning and feedback data	Domain examples, preference comparisons, corrected answers and user feedback used to adapt a model	Treat adaptation as a separate processing activity; service interactions do not automatically authorize reuse for training.
Validation and evaluation data	Records used to select models, measure performance or test safeguards	Prefer synthetic or appropriately approved records; isolate test data and protect sensitive examples and evaluation traces.
Grounding data	Evidence supplied to constrain or inform an answer, including databases, documents, search results and structured tool responses	Grounding is a function, not a privacy exemption. Assess purpose, accuracy, permissions, provenance and disclosure to model providers.
Retrieval-augmented generation or RAG	A process that retrieves selected information and supplies it to a model as context; a common grounding technique	Apply source permissions at retrieval time and at downstream disclosure. Include connectors, document chunks, indexes, metadata and caches in scope.

AI data layer	Meaning and examples	Data protection implications
Embeddings and vector stores	Numerical representations and indexes used for similarity search and retrieval	Evaluate identifiability and linkage risk; protect tenant boundaries, document identifiers and reconstruction paths. Do not assume embeddings are anonymous.
Inference inputs and working context	Prompts, uploaded files, conversation histories, system context and tool results used during a request	Minimize what enters context, constrain tools and destinations, and verify provider retention and reuse settings.
Outputs and operational records	Generated text, inferred attributes, scores, agent actions, audit logs, traces and support records	Review accuracy and disclosure; propagate rights and retention controls to derivatives; redact unnecessary logging.

RAG ordinarily supplies information at inference time rather than changing model weights. That can make source correction and access control more manageable, but it does not prevent leakage, secondary use or provider retention. A grounded answer may still be inaccurate, and an accurate answer may still disclose information the requester is not entitled to receive. Grounding and privacy assurance therefore require separate tests. These architectural distinctions guide the Institute's proposed assessment and control approach; they are not new legal categories.

3. Laws Regulations and Standards for AI Data Protection

AI applications can combine information gathered under different legal authorities, contractual commitments and confidentiality rules. A RAG assistant may retrieve personnel files alongside internal security procedures, while a learning pipeline may reuse customer records collected for an unrelated service. Regulation establishes enforceable boundaries for collection, use, disclosure, security and individual rights. It also creates accountability where affected individuals cannot inspect the system or meaningfully negotiate its data practices.

A cross-border deployment must determine which obligations apply to each processing activity and participant, rather than assuming that the hosting region alone decides the law. Sensitive personal information, confidential business information and officially classified information may coexist in the same workflow, but they are not interchangeable categories. The legal basis for personal-data processing and the authority to handle national-security information must each be established where relevant.

Sensitive and classified information. An organization's internal classification scheme, such as confidential or restricted, is distinct from formal government security classification. In the United States, classified national-security information is protected under specific legal authorities, while Controlled Unclassified Information (CUI) is unclassified information requiring specified safeguarding or dissemination controls. PII is not automatically classified information or CUI. Determine the applicable designation and handling authority before permitting ingestion, grounding, model training or tool access. The general controls in this article do not by themselves authorize classified processing: such use requires the relevant approved environment, access eligibility, need to know, dissemination controls and system authorization. Other countries use their own classification frameworks. Sources: NIST

classified-information glossary and NARA CUI guidance, listed below.

3.1 Distinguishing legal instruments and standards

Laws and statutes establish binding obligations under the applicable legal system. A statute is generally enacted by a legislature; the broader term law also encompasses other recognized sources of binding rules. Data protection legislation may establish rights, regulator powers, enforcement and sector-specific duties. An instrument's legal force depends on jurisdiction and authority, not simply its title.

Regulations are binding rules, but the term has different institutional meanings. In the United States, agencies generally issue regulations under delegated legal authority. In the European Union, a Regulation is a binding EU legal act that is directly applicable in Member States without national transposition; the GDPR is an example. National implementing measures and enforcement structures can still be necessary. Regulations should not be described universally as subordinate administrative guidance.

Directives also require contextual interpretation. An EU Directive binds Member States as to the result to be achieved and generally requires transposition into national law, leaving choices of form and method within the legal framework. An internal organizational or executive directive is a different kind of instruction, whose authority depends on its issuer and legal basis. Neither meaning is equivalent to a voluntary standard. The European Commission's explanation of types of EU law is referenced below.

Standards and frameworks provide specifications, management requirements, control guidance or assessment methods. ISO standards and NIST publications differ in purpose and applicability; a standard can become mandatory through legislation, regulation, procurement, contract or organizational adoption. Standards support implementation and assurance but do not supply a missing legal basis or replace every applicable legal obligation.

The regional comparison below should therefore be read as an applicability map. Establish the relevant legal duties first, then select and assess the controls needed to meet them. Record binding legal requirements separately from voluntary recommendations and internal policy choices.

NIST is a standards and guidance institution, not a general privacy regulator. Its publications can become binding through federal requirements, contracts or organizational adoption. ISO standards similarly differ from statutes: ISO/IEC 27001 specifies ISMS requirements; ISO/IEC 27002 provides control guidance. ISO/IEC 27701:2025 addresses privacy information management systems. Neither a control catalog nor certification replaces an applicability analysis.

Table 3

Region / jurisdiction	Main legal or governance layer	Practical significance for AI
United States	Federal sectoral laws and consumer protection, federal-agency privacy requirements, state privacy and security laws. Examples include HIPAA for covered activities, GLBA for relevant financial institutions, COPPA for covered children's services, and California CCPA as amended by CPRA.	Determine coverage rather than assuming one nationwide private-sector privacy rule. Evaluate collection, disclosure, profiling, retention, security and state-specific rights. HIPAA does not cover every health-related record or app.
U.S. federal systems	Privacy Act, applicable E-Government Act PIA duties and federal security requirements; NIST publications support implementation.	Determine authority, purpose, notices, system records, privacy assessment and security baseline obligations.
Canada	PIPEDA for applicable commercial activities; provincial private-sector, public-sector and health laws also matter.	Apply accountability, appropriate purposes, meaningful consent where required, limited use/retention, access and sensitivity-based safeguards. Check Quebec and other provincial duties separately.
Mexico	Federal Law on Protection of Personal Data Held by Private Parties, newly enacted March 2025; distinct public-sector regime.	Document notices, consent where needed, sensitive-data handling, security, transfers and ARCO rights. Do not assume the former regulator structure still applies.
EU / EEA	GDPR and relevant national and sectoral rules.	Establish Article 6 basis; Article 9 condition where applicable; minimization, rights, privacy by design, security, DPIA and transfer measures. Article 22 has specific decision and effect thresholds, not a prohibition on every automated recommendation.
United Kingdom	UK GDPR and Data Protection Act, with subsequent amendments.	Run a separate current-law review; do not assume EU amendments or timelines automatically apply in the UK. Detailed UK changes are outside this draft's verified scope.
Japan	APPI, PPC rules and guidance; additional rules for specified identifiers.	Define utilization purpose, apply security measures and supervision, assess third-party and overseas transfers, and support applicable individual rights.
Brazil	LGPD and ANPD rules.	Identify Article 7 or Article 11 basis, apply necessity/security/accountability, assess transfers and individual rights. Article 20 addresses review of decisions solely based on automated processing affecting interests; do not describe it as an unconditional statutory right to a human reviewer.

Region / jurisdiction	Main legal or governance layer	Practical significance for AI
Argentina	Law 25,326 and implementing rules.	Address data quality, consent and exceptions, security/confidentiality, individual rights and transfers.
Uruguay	Law 18,331 and implementing requirements.	Assess purpose, consent and exceptions, security, rights and international transfers.
Chile	Existing framework plus Law 21,719, scheduled to enter into force 1 December 2026.	As of this draft date the new regime is forthcoming, not already applicable. Prepare for expanded duties and automated-decision safeguards.

This South American selection illustrates different national regimes; it is not a single regional law or an exhaustive survey. Further editions should add Colombia, Peru and Ecuador with country-specific source review. Statutory sources and regulator guidance are listed below.

4. PII attributes and safeguarding requirements

4.1 Common attribute taxonomy

All rows below inherit a baseline of documented purpose and authority, minimum necessary collection, authenticated identities, least-privilege authorization, encryption in transit and at rest appropriate to risk, managed cryptographic keys, protected audit records, retention limits and tested deletion. Authentication establishes identity; authorization determines which records and operations that identity may access. Encryption cannot replace either.

Table 4

Attribute type	Definition / examples	Additional recommended protection	ISO/IEC 27002:2022 candidates	NIST SP 800-53 Rev. 5 candidates
Identity and contact	Name, address, phone, email	Tokenize where useful; mask support views; exclude unnecessary fields from prompts	5.12, 5.15-5.18, 5.34, 8.11, 8.24	AC-3, AC-6, IA-2, SC-8, SC-28, PT-2, PT-3
Government identifiers	Passport, tax and social-security identifiers	Restrict collection and full-value display; separate linkage keys; prevent model ingestion	5.34, 8.3, 8.11, 8.12, 8.24	AC-3, AC-6, SC-12, SC-28, SI-19, PT-7

Attribute type	Definition / examples	Additional recommended protection	ISO/IEC 27002:2022 candidates	NIST SP 800-53 Rev. 5 candidates
Authentication secrets	Passwords, recovery codes, access tokens linked to accounts	Password-specific salted adaptive hashing; vault recoverable secrets; rotate/revoke; keep out of prompts and logs	5.17, 8.5, 8.15, 8.24	IA-5, AC-6, AU-9, SC-12, SC-13
Financial and transaction	Account details, payment and purchase history	Payment tokenization; compartmentalization; transaction integrity and fraud monitoring	5.3, 8.3, 8.12, 8.15, 8.24	AC-5, AC-6, AU-9, SC-8, SC-28, SI-4
Health and genetic	Diagnoses, medications, genetic results and related predictions	Restrict use cases; approved processing environments; tight export controls and human validation	5.31, 5.34, 8.3, 8.12	AC-3, PT-2, PT-3, RA-3, SI-18
Biometric	Facial/voice/fingerprint templates and identifying measurements	Treat as difficult to replace; protect templates; evaluate necessity and spoofing; separate from general analytics	5.34, 8.3, 8.5, 8.24	IA-2, IA-5, AC-6, SC-28, RA-3
Location and behavioral	GPS, device IDs, browsing, movement and usage patterns	Reduce precision and retention; prevent cross-context linkage; control aggregation and queries	5.12, 5.34, 8.10, 8.11	PT-3, SI-12, SI-19, AC-4
Beliefs and affiliations	Religion, political views, ethnicity, union membership, sexuality	Limit inference and profiling; assess discrimination and disclosure harm; maintain exceptions register	5.31, 5.34, 8.3	PT-2, PT-3, PT-4, RA-3, AC-6
Children's and vulnerable-person data	Information about minors or people at heightened risk	Age/context safeguards; restricted sharing; heightened approval and assessment; avoid manipulative personalization	5.31, 5.34, 6.3, 8.26	PT-2, PT-4, RA-3, AT-2, SA-8
AI-derived and operational data	Scores, summaries, embeddings, prompts, transcripts, feedback and logs linked to people	Preserve provenance; enforce source permissions at retrieval; propagate correction/deletion; minimize logging	5.34, 8.10-8.12, 8.15, 8.26	AC-3, AC-4, SI-12, SI-18, SI-19, AU-9

Mapping status: these are selected candidate controls, not a complete baseline. ISO control identifiers are proposed cross-references requiring licensed-text validation before publication. NIST controls require tailoring, organization-defined parameters, applicable enhancements and assessment procedures. There is no blanket ISO/NIST rule that every PII field requires the same authentication factor or encryption algorithm. Select mechanisms against risk, applicable law and binding requirements. NIST Rev. 5 also has maintenance releases; record the adopted release in the implementation baseline.

4.2 North America: attribute interpretation and safeguards

Table 5

Attribute / context	Regional interpretation	Implementation overlay on section 4.1
U.S. health data	HIPAA protections depend on covered entity/business associate activities; other privacy and consumer laws may cover records outside HIPAA.	Establish role and disclosure authority; constrain AI vendor access and reuse; apply contractual safeguards where required.
California identifiers, precise location, genetics and credentials	CCPA distinguishes personal information and sensitive personal information; definitions and rights have statutory limits and exceptions.	Tag applicable categories, track purposes and sharing, implement applicable rights and retention limits.
Canadian personal information	Sensitivity depends on nature and context; even ordinarily non-sensitive data can become sensitive in combination.	Select safeguards proportionate to sensitivity and amount; maintain processor accountability and access/correction processes.
Mexican sensitive data	Includes intimate information and data whose misuse can cause discrimination or serious risk.	Verify consent rules and exceptions, ARCO workflow, notice and security measures; reassess overseas AI processing.

4.3 Europe: attribute interpretation and safeguards

Table 6

Attribute / context	GDPR distinction	Implementation overlay on section 4.1
Identifiers and linkable records	Personal data under Article 4; pseudonymization does not automatically remove scope.	Article 6 basis, purpose and retention controls; separate reidentification keys.
Health, genetics, religion, ethnicity and other Article 9 categories	Additional conditions apply; biometric data is an Article 9 category when processed to uniquely identify a person.	Record both Article 6 basis and Article 9 condition; tight access; DPIA where high-risk criteria are met.
Criminal-conviction/offence data	Governed separately by Article 10, not simply another Article 9 category.	Verify official-authority or authorized-law conditions and safeguards.
Profiles and consequential automated decisions	Articles 4 and 22; exact legal effects and involvement matter.	Review decision significance, exceptions, meaningful oversight and contestability; test actual human review.

Attribute / context	GDPR distinction	Implementation overlay on section 4.1
AI prompts, retrieval indexes and telemetry	Personal-data status depends on content and identifiability.	Include them in processing records, processor terms, rights searches, security design and transfer analysis.

4.4 Asia: Japan-focused attribute interpretation and safeguards

Table 7

Attribute / context	APPI distinction	Implementation overlay on section 4.1
Identifying descriptions and prescribed codes	Personal information; database-organized personal information may constitute personal data.	Specify utilization purpose; map storage and disclosure; apply appropriate security and staff/vendor supervision.
Medical history, creed, race and specified sensitive circumstances	Special care-required personal information; exact coverage is set by law and rules.	Check acquisition consent and exceptions; restrict use and transfers; do not copy GDPR category labels as if identical.
Individual identification codes	Includes prescribed identifying biometric conversions and assigned codes.	Verify whether the code is prescribed; minimize collection, protect templates and restrict linkage.
My Number	Subject to a separate statutory regime beyond ordinary APPI processing.	Establish purpose-specific legal review and strong segregation; exclude from general AI ingestion.
Pseudonymously / anonymously processed information	Distinct APPI categories with specific conditions; not interchangeable with ordinary masking.	Validate statutory processing and disclosure rules; test reidentification and do not assume GDPR equivalence.
Prompts submitted to cloud generative AI	PPC guidance emphasizes consistency with utilization purpose and checking that the provider does not use personal data for machine learning, among other considerations.	Verify actual service terms, configuration and processing roles; record reuse restrictions and monitor changes.

Japan is the focus of this Asian section, not a proxy for all Asian law. China PIPL, South Korea PIPA and India's DPDP regime require separate treatment in a later jurisdiction appendix.

5. AI regulation by regulatory type

Table 8

Regulatory type	Examples and status at draft date	Personal-information implications
Binding horizontal AI law	EU AI Act, with phased obligations and 2026 amendments reflected in Commission guidance	Classify roles and uses; assess prohibited practices, transparency and applicable high-risk duties. GDPR continues to apply independently.
Binding national AI governance framework	Japan's 2025 AI promotion law, fully effective September 2025	Government-led AI governance and risk-response framework; not a replacement for APPI or an equivalent copy of the EU risk-tier system.
Binding subnational AI / automated-decision law	Texas HB149 effective January 2026; Colorado's 2026 replacement ADMT law has substantive requirements scheduled January 2027 according to its Attorney General	Scope-specific restrictions, disclosures and decision protections; distinguish enacted future duties from currently applicable ones.
Binding general privacy law applied to AI	GDPR, APPI, LGPD, PIPEDA and relevant U.S. laws	Lawful processing, rights, security, purpose limits, vendor and transfer obligations apply regardless of an AI label.
Sector and use-specific law	Health, financial services, employment, children's services and biometrics	AI changes the processing mechanism, not the need to check existing sectoral duties.
Regulatory guidance and enforcement	PPC generative-AI guidance, privacy-regulator interpretations and enforcement	Clarifies how existing law applies; distinguish guidance from enacted statutory text and case-specific findings.
Voluntary standards and frameworks	NIST AI RMF and GenAI Profile; ISO/IEC 42001 and 23894; OECD principles	Useful governance and assurance resources; binding status may arise through contracts or adoption, not merely publication.

EU timing note: Commission pages reviewed for this draft report that prohibitions and AI literacy began applying in February 2025, GPAI obligations in August 2025, and most remaining rules including Article 50 transparency in August 2026. They report amended high-risk dates of 2 December 2027 for Annex III systems and 2 August 2028 for Annex I product systems. These dates must not be replaced with the superseded assumption that all high-risk duties applied in August 2026. Before publication, reconcile the guidance with the final consolidated legal text and transitional provisions for the particular system.

6. AI-specific risks and mitigating controls

The following scenarios are the Institute's analytical synthesis. A candidate mapping identifies relevant controls; the implementation and its evidence determine whether risk is reduced.

Table 9

Risk scenario	Safeguards to implement	ISO candidates	NIST candidates	Evidence to retain
Staff paste PII into unapproved AI services	Approved-service gateway, minimization, purpose checks, DLP and training	5.10, 5.23, 5.34, 6.3, 8.12	AC-20, AT-2, PT-3, SC-7	Policy, gateway rules and tested blocking outcomes
Prompt injection causes retrieval or tool disclosure	Treat retrieved content as untrusted; enforce authorization outside the model; least-privilege tools and bounded egress	8.3, 8.20, 8.26-8.29	AC-3, AC-4, AC-6, SC-7, SA-11	Cross-tenant tests, tool audit records and injection test results
Training memorizes rare personal records	Minimize/deduplicate data; use synthetic test records; evaluate extraction; consider validated privacy-preserving training	5.34, 8.11, 8.29	PT-3, SI-19, RA-3, SA-11	Dataset lineage and measured extraction results
Membership inference or reidentification	Evaluate attacker knowledge; limit query exposure; test deidentification; consider differential privacy with explicit utility/privacy tradeoffs	8.3, 8.11, 8.16	SI-19, AC-6, SI-4, RA-3	Attack evaluation and privacy parameters, where used
RAG exposes records beyond a user's permissions	Preserve source ACLs and enforce them at query time; tenant isolation; regression tests after permission changes	5.15-5.18, 8.3	AC-3, AC-4, AC-6	Positive and negative access tests, including stale-index cases
Inaccurate inferred health or employment claims cause harm	Provenance, correction process, uncertainty communication, meaningful review and challenge routes	5.34, 8.26, 8.29	SI-18, PT-5, RA-3, SA-11	Accuracy evaluations, appeals and correction records
Vendor retains prompts or trains on them unexpectedly	Assess processing roles, terms, settings, subprocessors, retention, location and reuse; monitor changes	5.19-5.23, 5.34	SA-9, SR-3, SR-6, PT-3	Contract, configuration evidence and supplier reassessment
Logs, traces and embeddings outlive source records	Field-level logging minimization; derivative inventory; lifecycle deletion and backup expiry	8.10, 8.11, 8.15	SI-12, AU-9, MP-6	End-to-end deletion tests and documented exceptions

Risk scenario	Safeguards to implement	ISO candidates	NIST candidates	Evidence to retain
Agent initiates unauthorized disclosure or record changes	Separate read/write authority; scoped credentials; approvals for consequential actions; transaction integrity	5.3, 8.2, 8.3, 8.15	AC-5, AC-6, IA-5, AU-12	Action authorization tests and attributable audit trail

A system prompt telling an assistant not to reveal PII is a behavioral layer, not an access-control boundary. Similarly, an output filter cannot make unlawful training lawful, and encryption at rest does not prevent an authorized inference endpoint from disclosing records.

6.1 Cryptographic controls across AI data flows

Cryptographic controls should be an explicit part of the AI data protection architecture. Select algorithms, protocols, implementations and validation requirements against applicable obligations and risk; do not assume that a vendor's statement that data is encrypted establishes adequate protection. The proposed measures below complement authorization, minimization and purpose controls.

Table 10

Protection objective	Recommended implementation	ISO 27002 candidates	NIST SP 800-53 Rev. 5 candidates
Confidentiality in transit	Use authenticated encrypted channels for users, ingestion connectors, model APIs, retrieval services, tools and inter-region transfers; validate peers and reject obsolete protocols.	8.20, 8.21, 8.24	SC-8, SC-13, SC-23
Confidentiality at rest	Encrypt source repositories, training/evaluation sets, vector stores, model checkpoints where sensitive, logs, exports and backups; assess temporary storage and snapshots.	8.13, 8.24	SC-28, SC-13
Key and secret lifecycle	Manage keys in approved KMS/HSM services where appropriate; separate key administration from data access; restrict service identities; rotate, revoke, recover and audit keys.	5.3, 5.17, 8.2, 8.24	SC-12, IA-5, AC-5, AC-6

Protection objective	Recommended implementation	ISO 27002 candidates	NIST SP 800-53 Rev. 5 candidates
Integrity and provenance	Use authenticated integrity mechanisms and signature verification where appropriate for datasets, models and releases; maintain trusted signing-key provenance. A bare checksum does not authenticate a source.	5.33, 8.24, 8.25	SI-7, SC-13, SA-10
Identifier separation	Tokenize or pseudonymize identifiers when useful; isolate reidentification keys and tightly authorize reversal. Encryption is not anonymization.	8.3, 8.11, 8.24	SI-19, AC-3, SC-28
Protected processing and retirement	Consider confidential computing where justified, with platform/attestation review; assess cryptographic erasure only where every relevant key and copy is controlled and destruction is verifiable.	8.10, 8.24	SC-39, SC-28, MP-6

Encryption at rest normally requires decryption for processing. It therefore does not prevent an authorized model service, compromised endpoint or overprivileged agent from accessing plaintext. Confidential computing can reduce particular infrastructure exposures but does not correct excessive application permissions or unlawful processing. Key recovery, ransomware resilience and availability must also be tested: inaccessible keys can make legitimate recovery impossible.

Evidence should include data-flow coverage, protocol and certificate configuration, key ownership and access policy, revocation tests, backup restoration tests, integrity checks and documented exceptions. Apply NIST-approved or validated cryptographic mechanisms where required by the governing environment; distinguish algorithm selection from implementation/module validation. These mappings remain candidate cross-references subject to the same validation gate as section 4.

7. Identifying and assessing PII risks

PII risk assessment should examine the relationship between the information an AI system uses, the actions it can perform and the effects it can have on people. The assessment boundary includes training and fine-tuning data, grounding sources, retrieval components, identities, tools, generated outputs and

operational records. Starting with the model alone misses many of the most consequential exposure paths.

Grounding data deserves particular attention because it can bring current, identifiable and permission-restricted records into an otherwise general-purpose model. The assessment must follow a record from its source through ingestion, indexing, retrieval and context assembly to the answer or action. It should establish whether the requester is entitled to receive the result, whether the processing purpose permits the use, and whether intermediaries retain or reuse the information.

Figure 1 illustrates the Institute's proposed assessment process. Data discovery and system mapping establish a shared evidence base. Privacy and legal analysis then run alongside security and misuse analysis. Their findings feed coordinated control design, testing and an AI risk register. Unresolved findings return to remediation; operational changes trigger renewed assessment. The process is iterative rather than a one-time approval checklist.

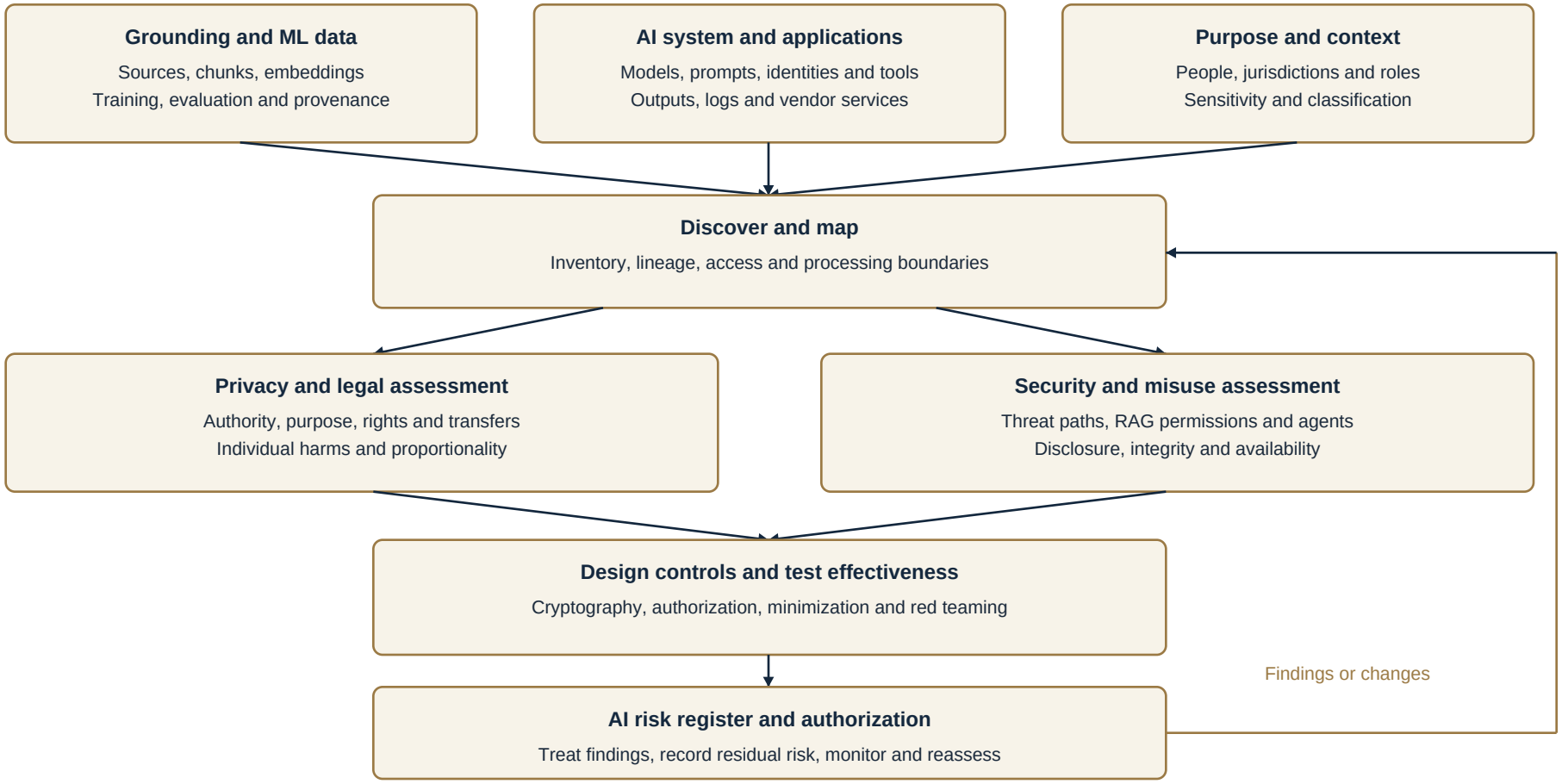


Figure 1. Integrated assessment of grounding data and AI systems. Parallel assessment tracks converge on tested safeguards and accountable authorization. The AI risk register records decisions and evidence; it does not itself make processing lawful.

7.1 Recommended assessment sequence

Organizations face an overlapping set of methods for identifying personal information and evaluating its risks: data discovery, classification, privacy impact assessment, business impact analysis, security risk assessment, threat modeling, penetration testing and AI red teaming. The methods answer different questions and produce different evidence. Selecting one as the organization's entire approach leaves predictable gaps.

The Institute recommends a coordinated assessment sequence organized around the AI use case and its data flows. First identify the information and processing purpose. Then evaluate the legal context and possible harms to people, trace technical misuse paths, test the proposed safeguards and authorize the residual risk. Assessments should share a common inventory and AI risk register so that findings do not disappear between privacy, security and business teams.

Automated discovery is useful for recognizable patterns such as account numbers, but it can miss personal information embedded in images, voice, free text, document metadata and inferred attributes. Combine discovery tools with data-owner interviews, schema and lineage review, representative sampling and manual assessment of linkability. Record both false positives and likely blind spots; a clean scan is not proof that a dataset contains no personal information.

A BIA principally examines disruption, dependencies and recovery. A PIA examines harms to individuals from processing, including processing that works as designed. Penetration testing examines exploitable weaknesses; AI privacy testing adds model- and retrieval-specific disclosure scenarios. The assessment plan should combine the methods proportionately, with deeper testing for sensitive data, consequential decisions, broad retrieval access, external tools and cross-border processing.

Table 11

Assessment	Main question	Required output / recommended timing
PII discovery and data-flow assessment	What personal information exists, where does it move, and who can use it?	Data inventory, lineage and trust-boundary map before procurement/design; refresh after changes
Privacy threshold screening	Does the use case require deeper assessment or special approval?	Recorded triggers and jurisdiction/role determination at intake
PIA / DPIA / applicable local impact assessment	Is processing justified and proportionate, and what risks does it create for people?	Alternatives, lawful grounds, rights analysis, mitigations and accountable decision before high-risk processing
Business impact analysis	What happens when the service or its data becomes unavailable or corrupted?	Dependencies, recovery objectives and continuity requirements
Security and privacy threat modeling	Who can misuse which flows, or be harmed without an adversary?	Abuse cases, trust boundaries and control requirements during architecture
Supplier and cross-border assessment	Who else processes data, for what purposes and under which transfer conditions?	Role allocation, terms, subprocessor map, transfer mechanism and risk evaluation

Assessment	Main question	Required output / recommended timing
Configuration review and penetration testing	Can identities, APIs, storage, networks or tenant boundaries be bypassed?	Authorized test findings, remediation and retest before release and after material change
AI privacy red teaming	Can prompts or tools extract data, infer membership or bypass retrieval boundaries?	Reproducible tests, success rates, severity and mitigations before release and periodically
Rights and deletion testing	Can access, correction, restriction and deletion operate across derivatives?	End-to-end evidence, limitations and exception handling
Control assessment and monitoring	Do safeguards work in operation?	Examine/interview/test evidence, monitored thresholds and reassessment triggers

NIST's PIA definition explicitly examines handling, privacy effects, legal conformity and alternative processes. SP 800-53A supplies control-assessment methodology; passing a vulnerability scan alone is not equivalent to this assessment.

7.2 AI risk register and approval criteria

Record asset/process, data subjects, attribute types, jurisdictions, purpose, processing role, threat or problematic action, individual harm, likelihood, severity, controls, evidence, owner, treatment and residual risk. Consider discrimination, exclusion, financial loss, reputational damage, physical safety, surveillance and loss of autonomy, alongside organizational impacts. Do not use a numerical risk score as a substitute for a prohibited-processing decision.

Recommended release gates are: documented processing authority and purpose; approved supplier/transfer arrangements; tested identity and retrieval boundaries; no untreated critical disclosure path; operational rights and incident workflows; and named acceptance of residual risks. Reassess when models, datasets, tools, access permissions, vendor terms or use cases materially change.

Use synthetic or consented test records wherever feasible, written test authorization and bounded environments. Avoid creating a new uncontrolled PII repository in the assessment evidence itself.

7.3 PII identification and classification in practice

Start by inventorying business processes, AI services, models, connectors, datasets and storage locations. Trace representative records from collection through training or inference to outputs, vendor systems, logs, backups and deletion. Include shadow AI use and spreadsheet/file exports rather than limiting the inventory to formally approved applications.

Discovery should combine structured pattern matching, contextual classifiers, optical character recognition for documents/images where appropriate, metadata inspection and manual sampling. Assess whether attributes become identifying when joined with public or internal reference data. Review model outputs for newly inferred sensitive attributes, not merely reproduction of input fields. Apply the legal categories of each relevant jurisdiction separately from the Institute's operational sensitivity rating.

For every material data flow, record the processing purpose, source, responsible owner, applicable roles, affected population, destination, retention rule and rights-handling path. Classify uncertainty explicitly: an unverified dataset should not silently receive a low-risk label because its contents are poorly understood. Discovery tools themselves require approved processing arrangements if they transmit sampled personal information externally.

7.4 Test design and evidence of risk reduction

Use a documented test plan linking each AI risk register scenario to a safeguard and an observable outcome. For a RAG assistant, test legitimate retrieval alongside denied access, cross-tenant queries, stale permissions, deleted source documents, malicious retrieved instructions and attempts to export results through tools. Confirm that source access revocation reaches indexes and caches within the approved interval.

For ML training or fine-tuning, inspect data provenance and sampling practices, then evaluate memorization and extraction using an explicitly defined attacker model. Membership-inference results require context: attack access, reference populations and baseline error rates affect their meaning. Test deidentification against realistic linkage opportunities. Keep test data, prompts, software/model versions and configuration sufficient to reproduce findings without unnecessarily retaining real PII.

For rights workflows, execute a representative access or correction request across source records, prompts, retrieval indexes and operational records. Test deletion and documented backup expiry. Identify model-weight limitations separately rather than treating deletion from a source database as proof of model-level erasure. Examine supplier evidence and conduct available technical checks; an assurance statement alone does not establish all downstream behavior.

7.5 Treatment priorities and reassessment

Treat an absence of processing authority, prohibited use or missing required transfer condition as a governance decision requiring resolution; technical safeguards do not cure it. Prioritize exploitable disclosure paths and high-severity harms, identify interim restrictions, and assign an owner and deadline to every accepted remediation. Residual risk acceptance should identify the decision maker, supporting evidence, limitations, expiry and review trigger.

Reassess after changes to models, training data, grounding sources, connectors, agent privileges, vendor terms, retention, geographic destinations or consequential decisions. Use ongoing sampling and targeted regression tests to complement periodic formal review. Aggregate findings across systems so repeated failures, such as stale RAG permissions or unredacted logs, lead to architectural fixes rather than isolated exceptions.

8. AI Data Protection and Risk Training

Training should translate governance into decisions people actually make. Staff must recognize personal information, minimize prompts, distinguish approved from unapproved services, understand provider reuse and retention, verify generated claims about people and report suspected disclosure promptly.

Institute training connection: I-AI-GCRM Individual/Home User AI Safety and Security series, IND-101-107. **Editorial placeholder: insert verified Education & Training URL and the relevant privacy module link.** A public destination was not verified for this draft; no URL has been invented. Confirm module titles and current versions against the authoritative training package before release.

Table 12

Audience	Learning outcome	Practical exercise / evidence
Individuals and general staff	Recognize PII; minimize prompts; choose approved services; report mistakes	Redact a sample prompt and identify what still permits linkage
Developers and data engineers	Implement data boundaries, retention, provenance and privacy tests	Build and test a retrieval permission boundary and deletion path
Security and privacy teams	Jointly assess lawful processing, threat paths and individual harm	Conduct a combined PIA and adversarial-test workshop
Procurement and legal	Allocate roles, reuse limits, subprocessors, transfer and incident duties	Review a sample AI vendor agreement against the actual architecture
Executives and system owners	Understand residual risk and authorization accountability	Decide a deployment case with explicit evidence gaps

Map awareness and role training to ISO 27002 control 6.3 and NIST AT-2/AT-3. Evaluate performance through exercises and reporting behavior, not attendance alone. Include EU AI literacy obligations where applicable to the organization's role and use.

9. Data Protection Accountability Incident Response and Lifecycle Assurance

An accountable system owner should coordinate with privacy, security, legal, procurement, engineering and affected business teams. Document controller/processor or other applicable roles rather than relying solely on a vendor's marketing designation. Contracts should address permitted purposes, model training/reuse, retention, deletion, assistance with rights, incident notification, subprocessors and evidence of safeguards.

Incident plans should cover prompt disclosures, cross-tenant retrieval, agent exfiltration, exposure through logs, memorized training information and harmful false statements about people. Establish containment, evidence preservation, notification analysis, vendor coordination and corrective action. Legal notification thresholds and deadlines differ; maintain a jurisdiction-specific incident matrix rather than applying one global deadline.

Deletion must cover source records, indexes, caches, logs and export copies, with backup expiry and legal holds documented. Removing a training row does not prove the model has forgotten it. Where training has occurred, assess retraining, unlearning or other remediation and validate their effectiveness; do not promise guaranteed erasure from model weights without evidence.

Useful measures include inventory coverage, percentage of systems assessed before launch, unresolved high-severity findings, cross-tenant retrieval test failures, deletion completion, rights-request completion, supplier reassessment and incident containment time. Interpret metrics with sampling limits and false-negative rates. A zero-incident count may indicate weak detection rather than effective protection.

10. Conclusion and research agenda

AI personal-information protection requires coordinated decisions about legitimacy, architecture, operations and assurance. Regional laws define different categories and duties, while ISO and NIST controls provide reusable implementation structures. The resulting program should trace every material processing purpose and risk to an owner, safeguard and testable evidence.

Further research should evaluate how reliably organizations propagate rights through retrieval systems and model artifacts, measure privacy leakage under realistic attacker conditions, and reconcile competing demands for transparency, auditability and data minimization. The next editorial version should add a worked case study and a fully validated control appendix using licensed ISO text and the adopted NIST catalog release.

References and verification notes

Sources below were located or reviewed for this working draft on 28 October 2026. Legal summaries are paraphrases. Primary sources take precedence over this article. A listed catalog page verifies scope/edition but does not establish that every licensed clause has been inspected. Standards and AI risk references marked supplemental require full-text review before publication.

• NIST. SP 800-122: Guide to Protecting the Confidentiality of PII; PII glossary.

• NIST. SP 800-53 Rev. 5, Security and Privacy Controls. Includes maintenance-release information; validate implementation against the adopted catalog.

• NIST. Privacy impact assessment glossary.

• NIST. SP 800-53A Rev. 5; AI RMF; GenAI Profile, AI 600-1. Supplemental full-text review pending.

• ISO. ISO/IEC 29100:2024; ISO/IEC 27002:2022; ISO/IEC 27701:2025. Licensed clause verification pending.

• ISO. ISO/IEC 42001:2023 and ISO/IEC 23894:2023. Supplemental AI management/risk references; licensed-text review pending.

• OECD. Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data.

• European Union. GDPR, Regulation (EU) 2016/679. Review Articles 4-6, 9-10, 12-22, 25, 28, 30, 32-36 and 44 onward for implementation.

• European Commission. AI Act overview and amended implementation timeline; AI Act Service Desk timeline. Consolidated-law reconciliation remains a publication gate.

• FTC. Consumer privacy business guidance. U.S. sectoral and state details require use-case-specific statutory verification.

• California. CCPA guidance; What is personal information?. Check current CPPA rules separately before implementation.

• Canada OPC. PIPEDA fair information principles; statutory text.

• Mexico. Federal private-sector personal data protection law, 20 March 2025.

• Japan PPC. Laws and policies, including consolidated APPI; generative-AI prompt guidance. Japanese originals control; translations aid understanding.

• Japan Cabinet Office. AI promotion law; English overview.

• Brazil ANPD. LGPD frequently asked questions; Brazil Legislature, LGPD Article 20.

• Argentina. Law 25,326, updated text.

• Uruguay. Law 18,331.

• Chile BCN. Law 21,719, effective 1 December 2026.

• Texas Legislature. HB149 enrolled summary.

• Colorado Attorney General. ADMT and chatbot rulemaking / current statutory transition.

• European Commission. Types of EU law.

• NIST. Classified information glossary.

• National Archives. About Controlled Unclassified Information.

Editorial gates for the next revision

• Validate ISO definitions and every ISO/NIST mapping against authoritative full text; identify the precise adopted NIST release and assessment procedures.

• Reconcile EU AI Act guidance and amended dates with consolidated legislation; expand detailed current U.S. sectoral/state and UK analysis.

• Verify training URLs, module titles and versions; preserve existing institutional branding when preparing a publication layout.

• Add a worked AI assistant case study, assessment template and evidence-based control test results. No empirical findings are claimed in this draft.

• Obtain jurisdiction-specific review of legal claims and applicability before public release. Publication of research version v0.3 authorized 29 October 2026; substantive revisions remain subject to review.

Institute disclosure and document status

This article is an Institute for AI Governance, Cybersecurity & Risk Management research version, prepared with AI-assisted research, drafting and document production. It presents educational analysis and proposed implementation approaches; it is not legal advice, a legal opinion, a compliance determination or a certification. AI-assisted material may contain errors or omissions and requires human editorial and subject-matter review.

Legal summaries reflect the stated review date and selected jurisdictions. Primary legal texts, current regulator guidance and applicable organizational obligations take precedence. Standards mappings are interpretive candidates, not official equivalence findings, and detailed ISO mappings require validation against licensed text. Referenced organizations and standards bodies have not endorsed this article. Sources are attributed to their respective publishers.

Status: published research version; publication authorized 29 October 2026. **Version:** v0.3, 28 October 2026. The Institute retains responsibility for publication decisions and final verification. The training link and source/control verification items identified in this draft remain open.