TECHNÉ AI

AI Governance Handbook

Regulation. Oversight. Evidence. Practice.

Khullani M. Abdullahi

September 7, 2026 · Edition 3.0.0

The complete free web edition, in print.
techne.ai/handbook/

Contents

Overview	3
Introduction	5
Legal & regulatory	7
ISO/IEC standards	8
EU AI Act	11
US federal	15
US states & cities	19
International	24
Sector-specific rules	28
Copyright & IP	32
Privacy, data & security	35
Technical safety	40
Frontier models	45
Guidance by role	49
Maturity planning	53
Glossary	61
Resources	66
Source guide	68
Edition history	70
About this publication	72

Links and chapter references open the corresponding online edition. The chapter footnotes preserve the source record. Keep this PDF's edition date with any excerpt or working note.

01 / OVERVIEW

AI Governance Handbook

techne.ai/handbook/

AI governance starts with a clear account of what a system does, who is responsible for it, and what evidence supports the decision to use it. This handbook helps you build that account across regulation, privacy, technical safety and organizational practice.

Free to read. No account required. This handbook is part of Techné AI's public research. It is separate from the paid working-document libraries and does not require a purchase or a form submission.

Choose a starting point

• **New to AI governance?** Begin with the Introduction, then use the Glossary as a working reference.

• **Responsible for legal or compliance review?** Start with Legal & Regulatory Frameworks. Identify the jurisdiction, role and activity before applying a requirement.

• **Building or buying an AI system?** Read Privacy, Data & Security and Technical Safety.

• **Preparing an executive or board discussion?** Use Guidance by Role and the illustrative Maturity Planning Frameworks.

• **Working with general-purpose or frontier models?** Read Frontier Models alongside the relevant legal chapter.

What changed in September

This edition updates regulatory coverage and corrects overbroad claims in the May edition. It distinguishes the enacted EU AI Act amendment from the earlier political agreement, replaces the superseded Colorado framework, adds New York's amended RAISE requirements, and clarifies federal, sectoral and international sources.

It also separates standards from certification guarantees, legal requirements from voluntary safety frameworks, and illustrative maturity stages from validated benchmarks. The web chapters and complete PDF are generated from the same reviewed content. See the edition history for the change record and limits.

From reading to a working record

Use the handbook to orient a review, then follow the material that fits the actual decision:

• Public briefings: deeper references on employment AI, board oversight, regulation and governance methods.

• AI Hiring Exposure Snapshot: a free starting point for an employment-AI discussion, not a compliance determination.

• TalentSight Intelligence: paid employment-AI working documents and library access.

• BoardSight Intelligence: board-oriented working documents, available by request.

- Contact Techné AI: discuss a defined advisory question or report an access problem.

Read the date as well as the rule

This is a dated educational reference, not an exhaustive register of every applicable law, a legal opinion, or a certification of any organization’s practices. Chapters identify sources and distinguish current provisions from proposals or historical examples. Check the underlying source for later changes, applicability and the full conditions before acting.

Found an error? Use the site’s corrections process. For the author’s background and publication scope, see About this publication.

02 / INTRODUCTION

Introduction

techne.ai/handbook/introduction/

AI governance connects an organization's decisions about AI with the people accountable for them, the controls used to manage risk, and the records that show what happened. It applies to systems an organization develops and to systems it buys, configures or uses through a supplier.

The aim is not to collect policies for their own sake. It is to make a practical decision explainable: what the system is intended to do, where it may fail, who may be affected, which requirements apply, and who can approve, change or stop its use.

How this edition is different

The September edition replaces the May edition's provisional or superseded descriptions where newer primary sources establish the position. In particular:

• The EU AI Act chapter distinguishes the adopted 2026 amendment, its substantive changes and separate transition dates. A delayed high-risk provision does not postpone every AI-related duty.

• The US state chapter replaces the former Colorado framework and includes New York RAISE. It separates ordinary AI use from duties that apply only to particular developers, deployments or decisions.

• The federal chapter distinguishes statutes, executive direction, agency procurement/use rules and voluntary NIST guidance. An executive order is not a blanket repeal of state law.

• The ISO chapter distinguishes management-system requirements, risk guidance, impact-assessment guidance and requirements for certification bodies. Certification is not proof that every system is safe or lawful.

• Copyright & IP separates the specific issues decided in selected cases from broader questions that remain fact-dependent or unresolved.

These are selected corrections and updates, not a claim that this publication covers every development. The source guide and each chapter's citations are part of the reading, not optional supporting decoration.

Five questions to answer first

1. **What is the actual system and use?** Record its intended purpose, users, affected people, data, outputs and decision process. A vendor's generic product description is not enough.

2. **What role does the organization perform?** Developer, provider, deployer, employer, service provider and public authority are not interchangeable categories. Legal definitions differ.

3. **Where does the activity occur or have relevant effects?** Consider the provision's territorial and subject-matter scope instead of assuming a single headquarters address settles applicability.

4. **What evidence is available?** Separate a claim, a contract promise, a test result and an operating record. Document material limitations and unanswered questions.

5. **Who owns the decision?** Name the person who can approve, restrict, escalate or stop the use. Set a review trigger when the system, purpose, data or legal position changes.

How to use the chapters

For a specific question, use the chapter map rather than reading every page first. Start with the relevant legal scope, then connect it to data/security controls and technical evaluation. The role guide turns those topics into responsibilities.

For program planning, use the six maturity frameworks as discussion prompts. They are illustrative, author-created descriptions—not a validated score, an industry benchmark, or a route to a guaranteed certification. Evidence can be uneven across dimensions, and a higher numbered stage is not necessarily the right next investment for every organization.

For a deeper employment or board question, follow the related Techné AI briefings. The handbook explains the landscape; the paid libraries offer separate working-document collections. Reading or downloading this free publication does not grant access to those libraries.

Keep different kinds of statements separate

Law: identify the provision, covered role, operative date, exceptions and enforcement context. **Guidance:** identify who issued it and whether it is voluntary or incorporated into a particular obligation. **A recommended practice:** identify the problem it is intended to address and test whether it works in the actual setting. **An example or forecast:** do not present it as a requirement or a measured result.

A useful governance record makes these distinctions visible. It also records uncertainty rather than silently turning an unanswered question into a favorable conclusion.

Looking forward

Build a process that can absorb change: maintain a system inventory, keep a source and decision log, name review owners, and define escalation and re-evaluation triggers. Revisit material changes in models, data, suppliers, users and law.

Those practices support better decisions; they do not eliminate risk or establish compliance by themselves. The practical test is whether people can use the record to understand, challenge and revise an actual AI decision.

03 / LEGAL & REGULATORY

Legal & Regulatory Frameworks

techne.ai/handbook/legal/

AI is governed by a combination of AI-specific provisions and existing rules on matters such as discrimination, privacy, consumer protection, product safety and intellectual property. The applicable combination depends on the activity, jurisdiction and organization's role. Standards and voluntary frameworks can organize the work, but are not interchangeable with law.

Find the relevant source

• International Standards (ISO/IEC): management-system requirements, risk and impact-assessment guidance, terminology, and requirements for certification bodies.

• EU AI Act: territorial and role-based scope, risk classes, GPAI obligations, transparency and the amended timetable.

• US Federal: executive policy, federal agency use and procurement, selected statutes, NIST and CAISI.

• US States & Cities: selected Colorado, Texas, California, New York, Illinois and other provisions, including distinctions between existing duties and future operative dates.

• International Regulation: selected regimes and initiatives outside the EU and United States.

• Sector-specific Rules: healthcare, financial services and employment overlays.

• Copyright & IP: selected decisions, data provenance, disclosure duties and limits on generalizing from individual cases.

Build an applicability record

For each potentially relevant rule, record the source and review date, the covered activity and role, the facts supporting applicability, the operative date, exceptions, required evidence and a responsible owner. Distinguish a current obligation from a proposal or an internal policy choice.

When several regimes apply, compare them provision by provision. The most demanding rule in one dimension may not satisfy a different rule's notice, assessment, recordkeeping or reporting requirements. A common control can support several obligations only after those differences are checked.

From scope to implementation

Connect the legal review to the system and data record, technical evaluation and named decision owners. For more detailed examples, use the site's multi-jurisdiction reference, employment references and board oversight material.

This handbook is a starting point, not an exhaustive applicability analysis. Consult the full source and qualified advisers where the decision requires it.

04 / ISO/IEC STANDARDS

International Standards (ISO/IEC)

techne.ai/handbook/legal/iso/

ISO/IEC standards serve different purposes: management-system requirements, risk and impact guidance, terminology, and requirements for certification bodies. **These roles are complementary, not interchangeable.** Two additions in 2025 were ISO/IEC 42005, published in May, and ISO/IEC 42006, published in July.[1][2]

This overview uses ISO's public descriptions and publication records. It does not reproduce or claim clause-by-clause verification of the licensed standards. Obtain the applicable edition for implementation, and separately determine legal and contractual requirements.

ISO/IEC 42001:2023 — AI management systems

Published in December 2023, **ISO/IEC 42001** specifies requirements for establishing, operating and improving an AI management system (AIMS). It is relevant to organisations developing, providing or using AI across sectors and sizes.[3]

The management system brings policies, responsibilities, processes and improvement activity into a defined organisational scope. A third party can assess conformity to the standard, but the certificate concerns that management-system scope. It should not be presented as proof that every model output is accurate, every AI product is safe, or every regulatory duty has been met.

For procurement, examine the certificate's legal entity, activities, systems, sites, validity and issuing body. A certificate covering one team or service does not establish coverage of a supplier's entire portfolio. Ask for supporting evidence relevant to the actual service you are buying.

ISO/IEC 23894:2023 — AI risk management

ISO/IEC 23894, published in February 2023, provides guidance on AI-related risk management and its integration into organisational activities. It is a guidance standard, distinct from the certifiable management-system requirements of 42001.[4]

An implementation can use 23894 to structure risk identification, analysis, treatment and review inside an AIMS. The practical task is to record context, affected people, uncertainty, controls and decision owners—not merely to list the standard in a policy.

ISO/IEC 22989:2022 — AI concepts and terminology

ISO/IEC 22989, published in July 2022, establishes AI terminology and concepts to support communication among different stakeholders.[5] Use a controlled glossary in contracts, inventories and incident records, while recognising that a statute may define the same term differently.

The handbook's Glossary is explanatory material, not a reproduction of this standard or a claim that every definition has been checked against it.

ISO/IEC 42005:2025 — AI system impact assessment

ISO/IEC 42005, published in May 2025, guides assessment of an AI system's foreseeable effects on individuals, groups and society throughout its lifecycle. It complements management-system and risk-management work.[1]

An organisation may use this guidance to organise an impact-assessment method. It is not a mandatory implementation recipe for every 42001-conforming organisation. Nor does using it automatically satisfy an EU AI Act Article 27 fundamental-rights assessment or a privacy impact assessment: scope, responsible party, required content, consultation and submission obligations need their own legal mapping. See EU AI Act.

ISO/IEC 42006:2025 — AIMS certification bodies

ISO/IEC 42006, published in July 2025, adds AI-specific requirements for bodies auditing and certifying an AIMS against 42001. It supplements **ISO/IEC 17021-1**; it is not the organisation's AIMS certification standard.[2]

The standard supports consistent, competent assessment. Certificates from different providers still need to be evaluated for their particular scope, validity and issuing body's status; publication of a standard does not establish those facts for an individual certificate.

Certification and accreditation are different. An external certification body issues a certificate; an accreditation body recognises a certification body's competence for a defined scope. ISO itself neither certifies organisations nor issues their certificates. Check the certification body's relevant accreditation and the certificate's current status rather than describing an organisation as "42006-accredited."[6]

Other relevant ISO/IEC references

Reference	Role
ISO/IEC 38507:2022	Guidance for governing bodies on the organisational implications of AI use
ISO/IEC 5338:2023	AI system lifecycle processes
ISO/IEC TR 24028:2020	Overview of AI trustworthiness; not a specification of required trustworthiness levels
ISO/IEC TR 24368:2022	Overview of ethical and societal concerns

A practical implementation sequence

The following is a suggested workflow, not a requirement to adopt every standard:

1. Define the organisational scope, AI uses, applicable laws and contractual promises.

2. Establish shared terminology, retaining any distinct legal definitions.

3. Build governance responsibilities and management processes; use 42001 where an AIMS is appropriate.

4. Select risk and impact methods suited to the context, drawing on 23894 and 42005 as useful.

5. Test whether controls operate in practice and retain evidence of decisions, monitoring and corrective action.

1. If certification serves a real customer or organisational need, agree the assessment scope and verify the external body’s competence and accreditation.

Existing quality or information-security processes may be reusable, but an ISO 9001 or ISO/IEC 27001 certificate does not establish AI management-system conformity. Likewise, a management-system certificate is not a substitute for legal analysis, product testing or ongoing oversight.

Footnotes

1. ISO. ISO/IEC 42005:2025 — AI system impact assessment.
2. ISO. ISO/IEC 42006:2025 — Requirements for AIMS audit and certification bodies.
3. ISO. ISO/IEC 42001:2023 — AI management systems.
4. ISO. ISO/IEC 23894:2023 — Guidance on AI risk management.
5. ISO. ISO/IEC 22989:2022 — AI concepts and terminology.
6. ISO. Certification, accreditation and checking certificate status.

05 / EU AI ACT

EU AI Act

techne.ai/handbook/legal/eu-ai-act/

The EU AI Act, **Regulation (EU) 2024/1689**, entered into force on **1 August 2024** and applies in stages. Its requirements depend on the system, intended use, organisation's role, and applicable exceptions. A general-purpose AI model and a downstream system built with it can have different obligations.[1]

The **Digital Omnibus on AI is enacted**, not pending. Regulation (EU) 2026/1744 was published on 24 July 2026 and entered into force on **27 July 2026**. It changes selected high-risk deadlines and other provisions; it does not postpone the entire Act.[2][3]

Risk classification

The familiar four-level summary is a reading aid, not four mutually exclusive legal compartments. Article 50 transparency duties can overlap with high-risk duties, and obligations such as AI literacy can apply outside the high-risk category.[1]

Category	What to assess	Illustrative uses
Prohibited practices	Exact Article 5 conditions and exceptions	Specified social scoring, harmful manipulation, untargeted facial-image scraping, and workplace or education emotion inference, subject to the statutory exceptions
High-risk systems	Article 6, Annex III and the relevant product legislation; assess exclusions rather than classifying by sector alone	Certain employment, education, credit, insurance, biometric, infrastructure and regulated-product uses
Transparency duties	The relevant provider or deployer obligation in Article 50	Human interaction with AI, synthetic-content marking, deepfakes, emotion recognition and certain public-interest text
Other systems	Whether particular AI Act provisions or other laws still apply	A lower-risk use is not automatically exempt from privacy, consumer, employment or sectoral law

The new prohibitions concerning specified AI systems for child sexual abuse material and non-consensual intimate content apply from **2 December 2026**. Their provider/deployer conditions and safeguards matter: they are not a blanket prohibition on every general-purpose content generator.[3]

Phased timeline

Date	Milestone	Position at 7 September 2026
1 August 2024	Original Regulation enters into force	Past
2 February 2025	Original prohibited-practice and AI-literacy provisions apply	Applicable; Article 4 has since been amended
10 July 2025	Voluntary GPAI Code of Practice published	Available
2 August 2025	GPAI rules and specified governance provisions apply	Applicable, with the existing-model transition below
27 July 2026	Digital Omnibus enters into force	Enacted
2 August 2026	General application date, including Article 50 transparency and Commission GPAI enforcement powers	Applicable, subject to scoped transitions
2 December 2026	Specified new Article 5 prohibitions; Article 50(2) deadline for relevant systems already marketed before 2 August 2026	Future
2 August 2027	Compliance deadline for GPAI models placed on the market before 2 August 2025	Future
2 December 2027	Specified Chapter III Sections 1–3 provisions for Article 6(2)/Annex III systems	Future; original date was **2 August 2026**
2 August 2028	Corresponding provisions for Article 6(1) regulated-product systems	Future; original date was 2 August 2027

The high-risk amendment expressly excepts Article 6(5) from the deferral. Article 111 also contains existing-system transition rules. Establish which provision and system group a date applies to before using it as a launch deadline.[1][3][2]

What applies now

Prohibited practices. The original Article 5 restrictions have applied since February 2025. Social scoring is not confined to government use. Each prohibition has its own elements; for example, the workplace/education emotion-inference restriction includes medical or safety exceptions. Screen the intended and reasonably foreseeable use against the text.[1]

AI literacy. Amended Article 4 requires providers and deployers to take measures supporting the development of AI literacy among relevant staff and others operating systems on their behalf, accounting for their knowledge, experience and context. It expressly does not require guaranteeing any particular individual's literacy level. Record role-appropriate learning and support.[3]

GPAI. Article 53 addresses model documentation, downstream information, copyright policy and a public training-content summary. Article 55 adds duties for models with systemic risk. Check applicable exemptions and the **2 August 2027** transition for models marketed before 2 August 2025; do not assume every model became subject to every duty in August 2025.[1]

Transparency. Article 50 has applied since **2 August 2026**. Providers of interactive systems must address AI-identity disclosure; providers of synthetic-content systems must address machine-readable marking. Deployers have distinct disclosure duties for deepfakes, certain public-interest text, emotion recognition and biometric categorisation. Exceptions and allocation of responsibility differ by paragraph.[4]

The Article 50(2) extension to **2 December 2026** concerns relevant systems placed on the market **before 2 August 2026**. It is not a general best-efforts exemption or a grace period for all Article 50 duties.[3]

The GPAI Code of Practice

The Commission describes the Code, published **10 July 2025**, as an adequate **voluntary means of demonstrating compliance**. Its Transparency and Copyright chapters address Article 53; Safety and Security addresses Article 55 systemic-risk duties.[5]

The official signatory list includes Google, Microsoft, OpenAI and Anthropic. xAI signed the Safety and Security chapter only and must demonstrate transparency/copyright compliance through other adequate means. Check the current list rather than treating signature status as permanent.[5]

Signing is not an exemption, certification or automatic presumption of conformity. Actual adherence matters, and non-signatories can demonstrate compliance through other adequate means. The Omnibus expressly distinguishes codes from harmonised standards that can confer a presumption of conformity.[3]

See Frontier Models for model-level governance and the separate U.S. testing arrangements.

High-risk duties: separate provider and deployer roles

For an in-scope high-risk system, the requirements include lifecycle risk management, appropriate data governance, technical documentation, logging, instructions, human oversight, accuracy, robustness and cybersecurity (Articles 9–15). Providers must also address their Article 16 duties, quality management and the applicable conformity-assessment route. Not every assessment requires a third-party notified body.[1]

Registration is governed by Article 49, with different obligations and exceptions for providers and certain deployers; Article 71 establishes the database. It is not a universal registration rule for every AI system.[1]

Deployers have separate operational obligations under Article 26. Article 27's fundamental-rights impact assessment applies to specified deployers, including public-law bodies, private entities providing public services, and deployers of the specified creditworthiness and life/health-insurance systems. It is not an obligation on every high-risk provider or deployer; the Article 27 scope and exclusions must be checked.[1]

ISO/IEC 42001 and supporting standards can help organise governance evidence. They do not themselves establish EU AI Act conformity or replace the relevant legal assessment. See ISO standards.

Penalties and enforcement

Article 99 sets ceilings, not automatic fines:[1]

Specified violation	Monetary / turnover ceiling
Article 5 prohibited practices	EUR 35 million / 7%

Specified violation	Monetary / turnover ceiling
Listed operator and transparency obligations	EUR 15 million / 3%
Incorrect, incomplete or misleading information supplied to specified authorities	EUR 7.5 million / 1%

For undertakings, the higher applicable ceiling generally applies; Article 99 includes lower-of-the-two treatment for SMEs, including start-ups. The Omnibus extends specified proportional treatment to small mid-cap enterprises; do not generalise every SME concession to every organisation or violation. Article 101 provides a separate Commission fine regime for GPAI providers, with a ceiling of 3% of worldwide annual turnover.[13]

National competent authorities enforce within their assigned responsibilities; the AI Office supports Commission oversight, including GPAI, and the European AI Board supports coordination. Commission GPAI enforcement powers, including fines, became applicable on **2 August 2026**—not together with all the August 2025 provisions.[6]

Practical review checklist

• Inventory systems and models, intended uses, markets and your provider/deployer role.

• Screen against the original Article 5 restrictions and the specified new December 2026 prohibitions.

• Record measures supporting relevant staff's AI literacy under amended Article 4.

• Check each Article 50 duty now; document eligibility before relying on its narrow marking transition.

• For GPAI, identify model-market dates, exemptions and any systemic-risk classification.

• For high-risk systems, record the applicable Annex, assessment route, role-specific duties and transition date.

• Keep a provision-level evidence map; do not use an ISO certificate or Code signature as a substitute for compliance analysis.

This chapter is a dated overview, not a determination that a particular system is lawful or compliant. Use the original Regulation together with the amending text and applicable implementing measures.

Footnotes

1. Regulation (EU) 2024/1689, especially Articles 5–6, 9–17, 26–27, 43, 49–55, 99, 101, 111 and 113; read as amended.
2. European Commission. Regulatory framework on artificial intelligence.
3. Regulation (EU) 2026/1744, especially Article 1(5), (38)–(40), recital 42 and Article 4.
4. European Commission. (2026, July 20). Guidelines on transparency obligations for providers and deployers of certain AI systems.
5. European Commission. The General-Purpose AI Code of Practice and official signatory list.
6. European Commission. (2026, July 31). Commission starts enforcing AI Act rules and new transparency requirements on 2 August.

06 / US FEDERAL

US Federal

techne.ai/handbook/legal/us-federal/

Federal AI policy combines statutes, executive directions, agency requirements and voluntary technical frameworks. **Do not treat an executive policy announcement as a repeal of underlying law, or an agency's internal AI plan as a rule for every private organisation.**

This chapter distinguishes those instruments as of **7 September 2026**. It is not a complete inventory of federal rulemaking, current litigation or transaction-specific export controls.

Executive Orders 14148 and 14179

Executive Order 14110 was revoked on **20 January 2025** by the initial-rescissions order. **EO 14179**, signed **23 January 2025**, then directed a new AI Action Plan and review of actions taken under EO 14110.[12]

EO 14179 required agencies to identify inconsistent actions and, as appropriate and consistent with law, suspend, revise or rescind them—or propose doing so. It also directed OMB to revise its AI-use and acquisition memoranda. **It did not automatically erase every regulation, guidance document or statutory obligation associated with the prior policy.** Check the action and authority separately.

Federal agency use and procurement

On **3 April 2025**, OMB issued:

• **M-25-21**, replacing M-24-10: governance, transparency and risk-management requirements for agency AI use, including specified high-impact AI safeguards.[3]

• **M-25-22**, replacing M-24-18: acquisition guidance addressing competition, interoperability, data use, performance and risk in federal AI procurement.[4]

M-25-21's compliance-plan requirement applies to agencies within its scope, not only cabinet departments. Agencies must publish a compliance plan or a determination that they do not use and do not anticipate using covered AI, with recurring updates. An agency's published plan describes its own implementation; it does not by itself impose equivalent duties on the firms it regulates.[3]

M-26-04, issued **11 December 2025**, adds implementation guidance for the administration's Unbiased AI Principles in agency LLM procurement. It addresses contract terms and vendor information, with specified exclusions and scope rules; it is not a general standard governing all commercial LLMs. Its agency procurement-policy update deadline was **11 March 2026**.[5]

For a supplier, the practical question is which solicitation, contract terms, agency policy and legal authority apply to the proposed service—not whether the supplier has made a generic "federal AI compliant" claim.

America's AI Action Plan

Published **23 July 2025**, the Action Plan groups policy recommendations under innovation, infrastructure, and international diplomacy/security. It is a government policy roadmap, not a single directly applicable private-sector statute.[6]

The accompanying measures include the **American AI Exports Program**, which supports full-stack export packages. The export order expressly requires compliance with relevant export controls; it should not be summarised as simply relaxing or replacing them.[7] Other July measures address data-centre permitting and federal AI procurement. Their distinct scopes matter.

National framework and state-law challenges

EO 14365, signed **11 December 2025**, is titled **Ensuring a National Policy Framework for Artificial Intelligence**. It directs a DOJ task force to challenge selected state laws, Commerce evaluation of state laws, funding-related actions within legal authority, an FTC policy statement, and a legislative recommendation.[8]

The FCC direction is to **initiate a proceeding to determine whether** to adopt a federal reporting/disclosure standard after the specified Commerce evaluation—not a declaration that such a standard already exists. The order's listed child-safety, infrastructure and state-procurement exclusions constrain the **legislative recommendation**; they are not blanket exemptions from every section of the order.

On **20 March 2026**, the White House published a national AI **legislative framework** and called on Congress to enact it. That announcement is a recommendation, not itself an enacted preemption statute.[9]

Neither announcement is a sound basis for ignoring a state law. Determine whether an applicable statute, valid federal measure or actual court order changes the obligation. This chapter does not infer current case outcomes from litigation announcements.

June 2026: advanced AI and national security

EO 14409, signed **2 June 2026**, directs cybersecurity initiatives, benchmarking of covered frontier models and development of a **voluntary** framework for government access before release to trusted partners. It expressly disclaims authorising a mandatory government licence, preclearance or permit for developing or releasing models.[10]

NSPM-11, dated **5 June 2026**, addresses AI in the national-security enterprise, including adoption, assurance, accountability and acquisition policy. Its federal national-security context should not be presented as a general private-sector certification requirement.[11]

Deadlines directing officials to develop a programme do not prove that every implementing action has been completed.

Center for AI Standards and Innovation (CAISI)

NIST's **CAISI** conducts research, evaluations and standards work relating to AI capabilities and security.[12] On **5 May 2026**, NIST announced testing agreements with **Google DeepMind, Microsoft and xAI**, building on earlier partnerships that it said had been renegotiated to reflect current directives.[13]

These collaborations support predeployment evaluations and research. Participation is not a government certification that a model is safe, nor does the announcement establish that every version has completed testing. See Frontier Models.

NIST AI Risk Management Framework

AI RMF 1.0, released **26 January 2023**, remains a voluntary framework, and NIST states that a revision is underway. Do not label draft work or a concept note "AI RMF 2.0."[14]

The four interconnected functions offer a useful structure:

Function	Working question
Govern	Who owns the decisions, policies and escalation process?
Map	What is the use context, who may be affected and what could go wrong?
Measure	What evidence and evaluation methods characterise the risks?
Manage	Which risks require treatment, monitoring or a changed decision?

The functions are iterative rather than a mandatory one-way sequence. The companion Playbook offers selectable suggestions, not an exhaustive checklist or a guarantee of legal compliance.[15]

Profiles and work in progress

• **NIST AI 600-1, Generative AI Profile:** released **26 July 2024**. The March 2025 adversarial-machine-learning taxonomy is a separate NIST AI 100-2 publication, not an update to this profile.[14][16]

• **NIST IR 8596, Cyber AI Profile:** the official record identifies a **December 2025 preliminary draft**; the January 2026 comment period has closed. It is organised around CSF 2.0, not a replacement for the AI RMF.[17]

• **Critical-infrastructure AI RMF profile:** a **concept note** was released **7 April 2026**; distinguish that development stage from a final profile.[14]

• **Control Overlays for Securing AI Systems (COSAiS):** NIST's development project includes a January 2026 annotated discussion outline. These materials are not a single final, mandatory AI overlay.[18]

Record the exact publication, edition and draft status used in an assessment.

TAKE IT DOWN Act

The **TAKE IT DOWN Act**, signed **19 May 2025**, addresses specified non-consensual intimate depictions, including digital forgeries. Criminal liability depends on the statute's elements and exceptions; not every synthetic image is covered.[19]

Covered platforms had until **19 May 2026** to implement the notice/removal process. Following a valid request, they must remove the identified depiction as soon as possible and within **48 hours**, and make reasonable efforts to identify and remove known identical copies. FTC enforcement of the platform obligations began on **19 May 2026**.[19][20]

For an in-scope service, maintain an accessible reporting process, response ownership and evidence of handling.

Export controls: non-enforcement is not the same as repeal

On **13 May 2025**, BIS announced that it would not enforce the AI Diffusion Rule and planned to formally rescind and replace it. It also issued chip-related guidance.[21]

In its **12 May 2026** decision, GAO distinguished the announced non-enforcement policy from a completed rescission and found the policy subject to Congressional Review Act submission requirements. GAO recorded Commerce's explanation that formal rescission was not yet final.[22]

These dated records do not establish that a particular export is permitted today. Check current EAR text, licence requirements, item classification, destinations, end users and end uses before a transaction; do not reduce the remaining controls to "China and a handful of countries."

Employment and sectoral law still matters

Changing AI policy or withdrawing agency guidance does not repeal the underlying statutes. Covered employment decisions remain subject to laws such as Title VII and the ADEA; disability, credit-reporting, lending and other duties require their own applicability analysis.[23]

State and local rules also need separate review, including Illinois's **HB 3773 / Public Act 103-0804** employment-AI amendment and **New York City Local Law 144**. See US State Laws and Sectoral.

Footnotes

1. The White House. (2025, January 20). Initial rescissions of executive orders and actions.
2. The White House. (2025, January 23). EO 14179 — Removing Barriers to American Leadership in Artificial Intelligence.
3. OMB. (2025, April 3). M-25-21 — Accelerating Federal Use of AI through Innovation, Governance, and Public Trust.
4. OMB. (2025, April 3). M-25-22 — Driving Efficient Acquisition of Artificial Intelligence in Government.
5. OMB. (2025, December 11). M-26-04 — Increasing Public Trust in AI Through Unbiased AI Principles.
6. The White House. (2025, July). America's AI Action Plan.
7. The White House. (2025, July 23). Promoting the Export of the American AI Technology Stack.
8. The White House. (2025, December 11). EO 14365 — Ensuring a National Policy Framework for Artificial Intelligence.
9. The White House. (2026, March 20). National AI legislative framework announcement.
10. The White House. (2026, June 2). EO 14409 — Promoting Advanced AI Innovation and Security.
11. The White House. (2026, June 5). National Security Presidential Memorandum / NSPM-11.
12. NIST. Center for AI Standards and Innovation.
13. NIST. (2026, May 5). CAISI frontier AI testing agreements announcement.
14. NIST. AI Risk Management Framework and current revision status.
15. NIST. AI RMF Playbook.
16. NIST. AI technical publication register.
17. NIST. IR 8596 preliminary draft record.
18. NIST. Control Overlays for Securing AI Systems project.
19. Public Law 119-12 — TAKE IT DOWN Act, especially Sections 2–4.
20. FTC. (2026, May). TAKE IT DOWN Act enforcement starts.
21. BIS. (2025, May 13). AI Diffusion Rule non-enforcement and planned rescission announcement.
22. GAO. (2026, May 12). Decision B-337935.
23. EEOC. Title VII and ADEA.

07 / US STATES & CITIES

US State Laws

techne.ai/handbook/legal/us-states/

State and local AI requirements differ in their scope, operative dates, remedies, and enforcement arrangements. Some regulate a particular use, such as hiring or synthetic media; others impose duties on developers of frontier models. Enactment does not mean that every duty is already operative. Federal policy and litigation must also be checked for the particular provision at issue; an executive order is not a blanket repeal of state law. See US Federal.

This chapter covers selected state and local instruments, reviewed on 7 September 2026. It distinguishes enacted obligations from future operative dates and reported enforcement findings.

Colorado — SB 26-189 (automated decision-making technology)

SB 26-189, signed **14 May 2026**, repealed and reenacted the framework created by SB 24-205. The replacement's principal developer and deployer duties begin **1 January 2027**. The earlier June 2026 deadline and its annual impact-assessment and risk-management-program requirements should not be used as the current compliance checklist.[1]

The replacement addresses automated decision-making technology (**ADMT**) used to materially influence consequential decisions about education, employment, housing, financial or lending services, insurance, healthcare, or essential government services and public benefits.

Core obligations:

• **Developers:** provide technical documentation on intended uses, training-data categories, known limitations, appropriate use and human review; notify deployers of material updates.

• **Developers and deployers:** retain records necessary to demonstrate compliance for at least three years.

• **Deployers:** provide consumer notice, explain the ADMT's role following an adverse outcome within the statutory period, and support rights to correct factually incorrect personal data and request meaningful human review and reconsideration.

Enforcement. The Attorney General enforces the act under the Colorado Consumer Protection Act. Before 1 January 2030, a 60-day notice and opportunity to cure applies where cure is possible. The act creates no new private right of action; it also addresses allocation of fault in discrimination actions under existing law.[1]

Texas — HB 149 (Texas Responsible AI Governance Act, TRAIGA)

Effective **1 January 2026**, TRAIGA prohibits specified uses, including intentional incitement of self-harm, harm to others or criminal activity, certain government social scoring, and intentional unlawful discrimination. Disparate impact alone does not establish discriminatory intent under its discrimination provision. Disclosure duties cover government AI interactions with consumers and AI used in healthcare service or treatment, subject to the statute's conditions. The act also establishes a regulatory sandbox.[2]

The Attorney General has principal enforcement authority. Penalties distinguish curable violations or breaches of a cure statement (**$10,000–$12,000**), uncurable violations (**$80,000–$200,000**), and continuing violations (**$2,000–$40,000 per day**). The 60-day cure protection requires correction plus a written statement, supporting documentation and preventive policy changes. State licensing agencies may impose additional sanctions in the circumstances specified by §552.106.[2]

California

The following California statutes address different actors and activities; their obligations should be assessed separately.

SB 53 — Transparency in Frontier Artificial Intelligence Act

Signed **29 September 2025**, SB 53 distinguishes frontier developers from large frontier developers. Its requirements include:[3]

• **Large frontier developers:** write, implement and publish a frontier AI framework addressing catastrophic risks.

• **Frontier developers:** publish a transparency report before or concurrently with deployment of a new or substantially modified frontier model. Large developers include additional summaries of risk assessments, results and framework implementation.

• **Incident reporting:** report critical safety incidents to the California Office of Emergency Services within 15 days of discovery. Qualifying imminent risks of death or serious physical injury require disclosure within 24 hours to an appropriate authority with jurisdiction.

• **Whistleblower protections:** protect specified employee disclosures; large developers must also provide an internal reporting process.

Evidence used for other frontier-model frameworks may be relevant, but compliance with one jurisdiction does not establish compliance with another.

AB 2013 — Training Data Transparency

The initial documentation deadline was **1 January 2026**. Covered developers must publish training-data documentation on their website for specified GenAI systems, services and substantial modifications released on or after 1 January 2022 and made publicly available to Californians. Publication is also required before subsequent covered releases. The documentation includes dataset sources, types, licensing, personal information, protected intellectual property and other listed characteristics, subject to statutory exceptions.[4]

SB 942 — California AI Transparency Act

AB 853 moved the principal operative date to 2 August 2026. Covered providers must offer a free detection tool and the **option** of a manifest disclosure for image, video and audio content. Specified latent provenance disclosures are required, with technical-feasibility qualifications. This is not a requirement to visibly label every output.[5]

AB 853 adds duties for large online platforms and GenAI hosting platforms from **1 January 2027**, and specified capture-device duties from **1 January 2028**. Its scope and technical requirements must be checked independently from EU transparency requirements.[6]

Other California laws

California also legislated on election-related synthetic media and digital replicas. Enforcement of particular election provisions has been litigated: the Attorney General's advisory identifies court restrictions affecting AB 2655 and AB 2839. That historical advisory does not establish the present status of every provision or appeal; review current orders before relying on an election-law obligation.[7]

Utah — SB 226 (amending the Utah Artificial Intelligence Policy Act)

Effective **7 May 2025**, SB 226 establishes distinct disclosure rules:[8]

• **Consumer transactions:** a supplier using GenAI must identify it when the individual clearly asks whether the interaction is with AI or a human.

• **Regulated occupations:** prominent disclosure is required for high-risk AI interactions, verbally at the start of a verbal interaction and in writing before a written interaction.

• **Disclosure safe harbour:** conspicuous identification as AI, an AI assistant or not human at the outset and throughout the interaction can protect against enforcement of the disclosure provision.

The safe harbour is based on disclosure; it is not a general exemption for organisations overseen by a sector regulator.

Tennessee — ELVIS Act

The **Ensuring Likeness, Voice, and Image Security Act** (ELVIS Act), effective **1 July 2024**, adds voice to the personal rights protected by Tennessee law. It addresses specified unauthorised uses and tools for reproducing an individual's voice or likeness, subject to statutory conditions and exceptions. Voice licensing and consent deserve particular attention in entertainment and advertising.[9]

Illinois

• **HB 3773**, effective **1 January 2026**, prohibits discriminatory effects from AI used for specified employment purposes and the use of ZIP codes as proxies for protected classes. It requires notice of AI use for those purposes and assigns notice implementation details to Department of Human Rights rulemaking.[10]

• The **Artificial Intelligence Video Interview Act** separately addresses notice, explanation and consent before AI analysis of applicant-submitted video interviews, along with restrictions on sharing, deletion on request and specified demographic reporting.[11]

New York State — RAISE Act

New York's frontier-AI law was amended by **S 8828 / A 9449**, signed as Chapter 96 on **27 March 2026**. The current provisions take effect **1 January 2027**. A frontier model exceeds **10^{26}** training operations; a large frontier developer has more than **$500 million** in annual gross revenue together with affiliates.[12]

The law differentiates developer-wide obligations from large-developer framework and transparency duties. Critical-incident reporting to the designated office generally has a **72-hour** trigger under §1422; qualifying imminent death or serious-injury risks require disclosure to an appropriate authority within **24 hours**. Apply the statute's knowledge and determination conditions rather than treating these clocks as interchangeable.[13]

New York City — Local Law 144 (AEDT)

Local Law 144 applies to covered AEDTs used by employers and employment agencies for hiring or promotion. The tool must have an independent bias audit within **one year before use**, with the required summary made public. Required notice to covered NYC-resident candidates and employees must precede use by at least **ten business days**. Enforcement began **5 July 2023**.[14]

Reported enforcement findings: the Comptroller's **2 December 2025** audit found weaknesses in complaint handling and review. Against DCWP's identification of one issue, auditors identified at least **17 instances of potential noncompliance**. These were audit findings, not 17 adjudicated violations. The report alone does not establish a current enforcement campaign or the status of employer investigations.[15]

Managing the differences

The statutes do not share a single Colorado/Texas template. A practical register should identify:

1. The covered actor, technology, person and decision or output.

2. The operative date, notice, records, reporting and review duties.

3. The enforcing authority, available remedies, cure conditions and interaction with existing law.

Shared evidence can reduce duplicated work, but a review, disclosure or framework sufficient for one statute may leave another obligation unanswered.

Footnotes

1. Colorado General Assembly. SB 26-189: enacted summary and signed act.
2. Texas Legislature. HB 149 enacted text, Chapters 551–554.
3. California Legislature. SB 53, Chapter 138.
4. California Legislature. AB 2013, Civil Code §§3110–3111.
5. California Legislature. SB 942, especially Business and Professions Code §22757.3.
6. California Legislature. AB 853, Chapter 674.
7. California Attorney General. Legal advisory on existing California laws and AI, including litigation notes.
8. Utah Legislature. SB 226 enacted text.
9. Tennessee General Assembly. 2024 legislative abstracts: Public Chapter 588, ELVIS Act.
10. Illinois General Assembly. Public Act 103-0804, §2-102(L).
11. Illinois General Assembly. Artificial Intelligence Video Interview Act, 820 ILCS 42.
12. New York Senate. A 9449 / S 8828 amendment history; GBS §1420; GBS §1421.
13. New York Senate. GBS §1422.

14. New York City DCWP. Automated Employment Decision Tools requirements.
15. New York State Comptroller. Report 2024-N-6, 2 December 2025.

08 / INTERNATIONAL

International Regulation

techne.ai/handbook/legal/international/

International AI governance combines statutes, existing sector law, administrative programmes and voluntary guidance. These instruments differ in who they cover and what legal consequences follow. This chapter records selected developments reviewed on **7 September 2026**, distinguishing enacted obligations, proposals and international commitments.

Canada — the C-27 proposal and existing obligations

Bill C-27, which included the proposed Artificial Intelligence and Data Act (**AIDA**), did not become law in the parliamentary session that ended on **6 January 2025**. The parliamentary record shows it at House committee consideration. Its proposed obligations are not operative law, and its legislative history does not establish a date or final form for replacement legislation.[1]

Assess AI deployments under the privacy, human-rights, consumer and sector laws applicable to the organisation and activity. Federal public-sector automated decisions also have their own administrative framework: the **Directive on Automated Decision-Making** applies within its stated government scope, rather than to every private business using AI.[2]

United Kingdom — existing regulators and AI Growth Labs

The government's **21 October 2025** announcement described a proposed **AI Growth Lab** and a call for evidence on a cross-economy sandbox. It did not enact a comprehensive AI statute or establish that one would commence in the second half of 2026.[3]

The legal-services **advisory AI Growth Lab** subsequently launched as a programme helping participants navigate existing regulation. Government guidance updated **27 August 2026** says applications opened on 3 August and close on 27 September 2026. It expressly states that participation provides no regulatory approval, endorsement or exemption from legal obligations. Legal-services regulators and the Information Commissioner's Office provide coordinated support.[4]

For implementation, identify applicable data-protection, online-safety, professional and sector requirements first. Treat sandbox advice and policy proposals according to their stated status. The programme's current application timetable is not a timetable for a general AI Act.

South Korea — AI Basic Act

The **AI Basic Act** and its Enforcement Decree came into force on **22 January 2026**. MSIT's implementation announcement distinguishes transparency obligations, advanced-AI safety obligations, and responsibilities associated with **high-impact AI**.[5]

• **Transparency:** determine the notice and output-disclosure duties for the relevant high-impact or generative-AI service.

• **Advanced-AI safety:** the decree describes a separate combination of training compute, technological advancement and potential fundamental-rights impact.

• **High-impact AI:** assess the specified application area and risks, then the applicable operator responsibilities. Areas include recruitment, credit assessment, healthcare and education.

• **Implementation period:** MSIT announced a grace period of **at least one year**, generally deferring fact-finding investigations and penalties, with exceptional investigations for serious harms.

MSIT leads implementation of the AI Basic Act. Personal-data obligations also require assessment under the separate privacy regime. Consult current guidance for detailed applicability, including requirements affecting overseas operators; the implementation period does not erase other applicable law.

Japan — AI Promotion Act

Japan's **Act on Promotion of Research and Development, and Utilization of AI-related Technology** was enacted on **28 May 2025**, partly commenced on **4 June 2025**, and fully commenced on **1 September 2025**.[6]

The act establishes an **AI Strategic Headquarters**, chaired by the Prime Minister, and a national AI Basic Plan. It provides for research support, guidelines, information gathering, investigation of cases affecting rights and interests, and guidance or advice to businesses. Business operators have responsibilities to cooperate with government measures.

The statute does not establish an AI Act-style penalty schedule. That does not make the statute itself voluntary or displace existing privacy, consumer and sector law. Distinguish its legal institutional framework from the nonbinding guidance issued within it.[6]

China — AI-generated synthetic-content labels

The **Measures for the Labeling of AI-Generated Synthetic Content**, published **14 March 2025**, became effective **1 September 2025**. They apply to specified online information-service providers and build on China's algorithm-recommendation, deep-synthesis and generative-AI instruments.[7]

The central distinction is between **explicit labels**, perceptible to users, and **implicit labels**, embedded in file data or metadata. The measures cover generated or synthesised text, images, audio, video and virtual scenes. They prescribe different duties for generation services, distribution services, app-distribution platforms and users.

Article 9 permits specified provision of content without an explicit label following agreement on user responsibilities and retention of relevant logs for at least six months. Other provisions address downstream identification and prohibit malicious removal, alteration or concealment of labels. Applicability depends on the activity and conditions in the measures. The Chinese original is controlling.[7]

Brazil — PL 2338/2023 remains a proposal

The Chamber of Deputies' record for **PL 2338/2023** shows the proposal awaiting the rapporteur's opinion in the special committee at this review. Its **29 April 2025** entry records constitution of, and referral to, that committee; it is not a completed committee report.[8]

The Senate-origin proposal addresses AI development and use, risks and protections for individuals. Its text and timetable remain subject to the legislative process. An anticipated passage date is not an operative compliance deadline. Existing Brazilian law must be assessed independently while the proposal proceeds.

Singapore — governance and testing tools

Singapore's **AI Verify** toolkit and model governance frameworks provide technical and organisational resources. The **Model AI Governance Framework for Agentic AI**, launched **22 January 2026**, addresses bounded autonomy, meaningful human accountability, lifecycle controls and user responsibility. It complements earlier conventional-AI and GenAI guidance.[9]

IMDA subsequently announced an update incorporating additional case studies and practices. Use the current framework appropriate to the system, especially where an agent can take actions through tools. These resources support risk management; they do not establish universal safety or substitute for applicable law.[10]

OECD and UNESCO

The **OECD AI Principles**, adopted in 2019 and updated in **May 2024**, provide five values-based principles and policy recommendations. They address inclusive growth, human rights, transparency, robustness and accountability. They are a reference for policy and governance, not proof of compliance with every applicable law.[11]

UNESCO's Recommendation on the Ethics of Artificial Intelligence, adopted in **November 2021**, addresses human rights, dignity, fairness, human oversight and policy areas including data governance and education. Its assessment tools can help governments and organisations examine ethical implications; the recommendation should be distinguished from domestic legislation.[12]

Summit declarations and scientific reports

The **Paris AI Action Summit**, held **10–11 February 2025**, issued the Statement on Inclusive and Sustainable AI for People and the Planet. The presidency's published signatory list includes countries and regional organisations; the US and UK are absent from that list. Consult the current list when reporting participation, since a fixed count can become outdated.[13]

The **International AI Safety Report 2025** was published on **29 January 2025**, ahead of Paris. The **2026 report**, published **3 February 2026**, is the subsequent scientific assessment of general-purpose AI capabilities, risks and mitigation. Scientific reports and political declarations serve different purposes; neither by itself imposes domestic duties.[14]

India's AI Impact Summit has already taken place: the February 2026 New Delhi meeting concluded with the **New Delhi Declaration on AI Impact**. The Indian government's 21 February release records endorsements by countries and international organisations.[15]

Comparing instruments before applying them

Jurisdiction or instrument	Verified status or development	Practical distinction
Canada C-27/AIDA	Prior-session proposal did not become law	Assess existing law and the government directive separately
UK AI Growth Lab	Legal-services advisory programme open in August 2026	Participation does not waive law
South Korea	Act and decree effective 22 January 2026	Separate transparency, advanced-AI safety and high-impact duties
Japan	Full commencement 1 September 2025	Statutory institutions and cooperation duties; guidance is distinct
China	Labeling measures effective 1 September 2025	Actor, content type and explicit/implicit labels matter
Brazil PL 2338/2023	Awaiting special-committee opinion	Proposed text is not an operative statute
Singapore	Model frameworks and testing tools, including 2026 agentic guidance	Governance resources; legal duties remain separate
OECD, UNESCO and summit statements	International principles and commitments	Identify domestic implementation before treating them as law

For EU law, US federal policy and US state requirements, use the dedicated chapters and their dated source records.

Footnotes

1. Parliament of Canada. Bill C-27, 44th Parliament, first session.
2. Treasury Board of Canada. Guide on the Scope of the Directive on Automated Decision-Making.
3. UK Government. AI Growth Lab call for evidence.
4. UK Government. Legal services advisory AI Growth Lab: overview, updated 27 August 2026.
5. Ministry of Science and ICT. AI Basic Act implementation announcement.
6. Cabinet Office of Japan. Official outline of the AI Act; Ministry of Justice. Act translation.
7. Cyberspace Administration of China and other issuing authorities. AI-generated synthetic-content labeling measures.
8. Chamber of Deputies. PL 2338/2023 official legislative record.
9. IMDA. Launch of the Model AI Governance Framework for Agentic AI.
10. IMDA. Updated agentic framework and deployment initiatives.
11. OECD. AI Principles.
12. UNESCO. Recommendation on the Ethics of Artificial Intelligence.
13. Presidency of France. Paris statement and published signatory list.
14. International AI Safety Report. 2025 publication; 2026 publication.
15. Government of India, Press Information Bureau. New Delhi declaration and summit conclusion, 21 February 2026.

09 / SECTOR-SPECIFIC RULES

Sectoral Regulation

techne.ai/handbook/legal/sectoral/

An AI system can be subject to existing sector law even when no AI-specific statute applies. This chapter distinguishes binding requirements from agency guidance, voluntary frameworks, and proposals in selected US sectors. Status is reviewed as of **7 September 2026**. See International, US State Laws, and EU AI Act for additional jurisdictional overlays.

Healthcare — device classification and change control

FDA's current **Predetermined Change Control Plan (PCCP) final guidance is dated August 2025**, not December 2024. It provides nonbinding recommendations for marketing submissions involving AI-enabled device software functions. FDA reviews a proposed PCCP as part of the relevant marketing submission; changes implemented consistently with the authorised plan can avoid an additional marketing submission. This is not a blanket exemption for every model update.[1]

The guidance describes three components:

- **Description of Modifications** — the planned changes.
- **Modification Protocol** — the methods for developing, validating, and implementing them.
- **Impact Assessment** — the expected effects of those changes.

Classification comes first. FDA issued its current **Clinical Decision Support Software final guidance in January 2026**. Certain functions satisfy the statutory non-device exclusion; other decision-support functions remain subject to device regulation. A product being described as "clinical AI" does not resolve that question. FDA's January 2025 AI-device lifecycle and marketing-submission guidance is still listed as **draft**, which should not be described as a final rule.[2][3]

Other healthcare duties may apply independently of FDA classification. **45 CFR 92.210**, not 42 CFR Part 92, addresses covered entities' use of patient-care decision-support tools. Its text prohibits discrimination and requires reasonable efforts to identify specified risk factors and mitigate resulting discrimination risks. Applicability depends on the covered entity and activity; this is not a universal rule for every non-device AI product. Check relevant court orders and current HHS implementation before relying on a particular Section 1557 provision.[4]

Financial services

The OCC, Federal Reserve, and FDIC issued **revised interagency model-risk-management guidance on 17 April 2026**, replacing the 2011 guidance. OCC Bulletin **2026-13** expressly rescinds OCC Bulletin 2011-12 and other listed guidance. The old SR 11-7/OCC 2011-12 framework should therefore not be presented as the unchanged current instrument.[5]

The revised guidance:

- Addresses model development, use, validation, monitoring, governance, and third-party products.
- Is most relevant to banks with more than **$30 billion in assets**, although smaller banks with significant model-risk exposure may also find it relevant.

• **Excludes generative and agentic AI from its scope**. It is not an AI-specific supervisory standard covering every bank AI deployment.

• Does not establish enforceable standards or prescriptive requirements; the bulletin distinguishes supervisory guidance from law.

Separately, the CFPB's **26 September 2025 AI Compliance Plan** implements OMB M-25-21 for the **Bureau's own use of AI**. It is not a new supervisory framework governing lenders' AI in lending, servicing, and collections.[6]

Existing substantive law must be assessed on its own terms. For example, **ECOA and Regulation B** govern covered credit activity, including discrimination and adverse-action notification requirements. Using a complex model does not itself remove those obligations. Consumer-reporting, housing-lending, privacy, and other requirements need their own applicability analysis.[7]

Employment

Title VII, the ADA, and the ADEA can apply to employment tests and selection procedures, including computer-mediated tools. The relevant question is the effect and use of a procedure, not whether a vendor markets it as AI. EEOC's published discussion addresses discriminatory exclusion, job-related validation, and disability accommodation. A vendor's claims do not substitute for the employer's assessment of its own use.[8]

State and local duties are additional and differ in scope. See US State Laws for **NYC Local Law 144**, **Illinois HB 3773**, and Colorado's replacement law, **SB 26-189**, whose principal duties begin on **1 January 2027**.

Practical controls include:

• Determine whether a hiring or promotion tool is a **covered automated employment decision tool** under NYC's law before applying its bias audit, publication, and notice requirements. These requirements do not attach to every New York City hiring process.

• Evaluate job relevance, selection outcomes, and validation obligations under the applicable employment-law framework. Do not treat the Uniform Guidelines as a universal requirement to commission the same study for every tool.

• Provide a process to request reasonable accommodations and assess tools that may screen out qualified applicants with disabilities.

• Track candidate and employee notices separately for each applicable state or local law.

This chapter does not rely on an unverified date for the withdrawal of earlier agency AI guidance; changes in guidance should not be conflated with repeal of underlying statutes.

Consumer protection — FTC

Section 5 of the FTC Act prohibits unfair or deceptive acts or practices within the Commission's jurisdiction. AI marketing, product representations, and data practices therefore require analysis under existing consumer-protection law; an "AI" label neither establishes a violation nor creates an exemption.[9]

The FTC began enforcing the **TAKE IT DOWN Act's platform obligations on 19 May 2026**. Covered platforms must provide the prescribed notice-and-removal process for qualifying nonconsensual intimate imagery, including qualifying synthetic imagery, and comply with the **48-hour removal requirement after a valid request**, alongside the Act's other duties. This is not a general requirement to remove any allegedly harmful AI output within 48 hours.[10]

Useful governance measures include keeping substantiation for performance claims, checking whether actual data use matches representations, and assigning complaint and removal-request owners. NIST AI RMF can help organise that work; use of the framework is not a legal safe harbour.

Telecommunications

The FCC's **8 February 2024 declaratory ruling** confirms that AI technologies generating human voices fall within the TCPA's restrictions on artificial or prerecorded voices. The relevant calls generally require prior express consent unless an emergency-purpose exception or applicable exemption applies; telemarketing and other call-specific requirements must also be checked. The ruling is not a ban on every use of synthetic speech.[11]

The US Federal chapter addresses subsequent executive-branch directions concerning AI disclosure. A direction to consider or initiate rulemaking should not be represented as an already-operative FCC disclosure standard.

Critical infrastructure

NIST is **developing** an AI RMF Profile for Trustworthy AI in Critical Infrastructure. Its 2026 concept note describes a work programme, not a completed profile or a binding sector regulation.[12]

Operators should separately identify the actual cybersecurity, safety, reliability, procurement, and incident-reporting requirements that apply to their facilities and activities. Do not assume that every sector rule has acquired an AI-specific provision, or that applying a voluntary NIST profile satisfies those obligations.

Other sector overlays

• **Education** — FERPA restricts disclosure of personally identifiable information from education records at covered institutions. Vendor access must fit consent or an applicable exception and its conditions; use of an AI service is not itself an exception.[13]

• **Transportation** — NHTSA's Standing General Order requires specified manufacturers and operators to report certain crashes involving **automated driving systems or SAE Level 2 advanced driver-assistance systems**. It is not a reporting mandate for every vehicle containing AI.[14]

• **Defence** — DoD Directive **3000.09**, effective 25 January 2023, governs specified autonomous and semi-autonomous weapon systems, including human judgment, testing, and review requirements. Its scope and exceptions matter; it is not a general commercial AI law.[15]

• **Insurance** — NAIC adopted its **Model Bulletin on the Use of Artificial Intelligence Systems by Insurers** in December 2023. It describes expectations including risk-based written AI-system programmes and compliance with existing insurance law. The model bulletin is not self-executing nationwide: identify each relevant state's adoption, modifications, and supervisory requirements.[16]

How to navigate sectoral overlap

An organisation may face a product regime, sector requirements, employment or consumer-protection law, and state AI duties simultaneously. A useful compliance register should:

1. **Map applicability** for each system, legal entity, jurisdiction, and intended use.

2. **Separate authority types**: statute or regulation, supervisory guidance, draft, voluntary framework, and contractual requirement.

3. **Reuse evidence where appropriate**, while retaining regime-specific analyses, submissions, notices, and records. Conformity with a management-system standard does not establish compliance with every sector rule.

4. **Assign owners and review triggers** for model changes, new uses, incidents, and changes in law.

5. **Seek specialist or regulator input where appropriate**, particularly when device classification, a required submission, or an ambiguous supervisory scope affects deployment.

Footnotes

1. FDA. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions (final guidance, August 2025).
2. FDA. Clinical Decision Support Software (final guidance, January 2026).
3. FDA. Guidances with Digital Health Content.
4. eCFR. 45 CFR 92.210: Nondiscrimination in the use of patient care decision support tools.
5. OCC. Bulletin 2026-13: Revised Guidance on Model Risk Management (17 April 2026).
6. CFPB. AI Compliance Plan for OMB M-25-21 (26 September 2025); Artificial Intelligence at the CFPB.
7. CFPB. 12 CFR Part 1002: Equal Credit Opportunity Act (Regulation B).
8. EEOC. Employment Tests and Selection Procedures.
9. FTC. Federal Trade Commission Act.
10. FTC. FTC Begins Enforcing the TAKE IT DOWN Act (19 May 2026).
11. FCC. Declaratory Ruling FCC 24-17 (8 February 2024).
12. NIST. Concept Note: AI RMF Profile for Trustworthy AI in Critical Infrastructure.
13. US Department of Education. Privacy and Data Sharing.
14. NHTSA. Standing General Order on Crash Reporting.
15. DoD. Directive 3000.09: Autonomy in Weapon Systems (25 January 2023).
16. NAIC. Members Approve Model Bulletin on Use of AI by Insurers (4 December 2023).

10 / COPYRIGHT & IP

Copyright & IP

techne.ai/handbook/legal/copyright/

AI copyright governance involves several different acts: acquiring material, retaining a corpus, making training copies, distributing a model, and generating or publishing outputs. A favourable ruling about one act does not clear the others. This chapter distinguishes selected verified decisions from practical risk controls; it is not a complete docket tracker or a licence to use a particular dataset.

Bartz v. Anthropic: separate training from the library

In his **23 June 2025** partial-summary-judgment order, Judge William Alsup held that the training use at issue was fair use. He separately accepted the one-for-one digitisation of purchased print books for Anthropic's library, while rejecting fair use for downloading and retaining pirated books to build a general-purpose central library. The order did **not** hold that all training on pirated copies was infringing; nor did the training holding excuse the separate library acquisition. Its conclusion expressly left other uses of library copies outside the training ruling.[1]

The governance takeaway is to document **each acquisition and use**, rather than treating "AI training" as a single permission question. Preserve evidence of source, acquisition method, licence terms, allowed uses and deletion decisions. This is a practical inference from the decision, not a new statutory documentation duty.

On **20 July 2026**, Judge Araceli Martínez-Olguín granted final approval of the **$1.5 billion** class settlement and entered judgment. The order identifies **482,460 works** on the Works List and explains the limited release; future misconduct and model-output claims are not released. The often-repeated "about $3,000 per book" figure is not a guaranteed net payment to each author: fees, allocation between rightsholders and the approved distribution process matter.[2]

A district-court ruling and a class settlement are not a nationwide appellate determination of every AI-training practice. Settlement administration is also distinct from the merits of an unrelated copyright claim.

Kadrey v. Meta: a merits ruling on a limited record

Judge Vince Chhabria's **25 June 2025** order granted Meta partial summary judgment on its **fair-use defence** to the training-copying claim. That is a substantive fair-use ruling, not merely a procedural dismissal. The court stressed that these plaintiffs had not developed the market-harm record needed for their claims, especially the potential for market dilution by AI-generated competing works. The result was expressly limited to the record before the court; it was not a blanket approval of Meta's training practices or a general permission to acquire pirated material.[3]

For governance teams, market effects, the actual use, the relevant works and the evidence all matter. "Another developer won" is not a substitute for analysis of your own facts.

Getty Images v. Stability AI: a UK judgment, not a global training clearance

The UK High Court issued its judgment on **4 November 2025**, rather than merely commencing a trial in 2025. The training claims were no longer pursued on the territorial evidence before the court; the remaining secondary-copyright claim failed, while limited trade-mark claims concerning generated watermarks succeeded. These distinct outcomes do not decide the legality of training worldwide. The judgment is a useful reminder to evaluate outputs and branding separately from training-data rights.[4]

Other litigation concerning text, images, music and copyright-management information continues to require case-specific monitoring. This edition does not assert a current procedural status for every *New York Times*, *Andersen*, *Concord* or other action without a separately checked docket. Complaints are allegations, and interlocutory orders are not final liability judgments.

EU AI Act copyright obligations

For in-scope GPAI providers, **Article 53(1)(c)** addresses a policy to comply with Union copyright law, including appropriately reserved text-and-data-mining rights under Article 4(3) of Directive (EU) 2019/790. **Article 53(1)(d)** addresses the sufficiently detailed public training-content summary. The letters are different and should not be reversed. Applicable dates and transitional treatment depend on the model; see EU AI Act.[5]

The **GPAI Code of Practice** is a voluntary compliance tool. Its Copyright chapter addresses a copyright policy, lawful access, machine-readable reservations, output-related safeguards and a complaints contact. A signature alone does not establish implementation. Neither an ai.txt file nor a commercial opt-out registry is a universal statutory safe harbour; evaluate the actual reservation, relevant technical standard and applicable law.[6]

California training-data transparency

AB 2013 requires covered developers, beginning **1 January 2026**, to post specified training-data documentation before making covered generative AI systems or services, or substantial modifications, available to Californians. Coverage concerns releases on or after **1 January 2022** and includes statutory exclusions. The required fields include sources, dataset size and characteristics, copyright/public-domain status, personal information and acquisition details. Publishing a summary does not itself license the data or resolve infringement.[7]

Outputs and human authorship

The US Copyright Office's **Part 2 report was published on 29 January 2025**, not March. It explains that copyright protects human-authored expression: using AI as a tool does not disqualify an otherwise protectable work, but purely AI-generated elements are not protected through human prompting alone. Human selection, arrangement or modification can be relevant, depending on the facts. The report addresses copyrightability, which is a different question from whether an output infringes someone else's rights.[8]

Questions about memorisation, substantial similarity, copyright-management information, contractual rights and cross-border conduct must still be analysed separately. Record which elements people created and edited, and retain the permissions needed for source material and publication.

Practical controls

1. **Inventory sources and rights.** Record provenance, acquisition, licence scope, restrictions and accountable owner for each corpus.

2. **Escalate contested material.** Do not treat availability on the web as permission; isolate material whose provenance or rights cannot be established.

3. **Separate uses.** Review acquisition, retention, training, retrieval, fine-tuning and outputs independently.

4. **Implement applicable reservations and disclosures.** Preserve evidence of your processes and publish required summaries for covered offerings.

5. **Provide a rights contact.** Triage complaints, preserve relevant records and define correction or removal procedures.

6. **Test outputs.** Evaluate memorisation, close copying and unwanted marks; filters reduce risk but do not guarantee non-infringement.

7. **Review contracts.** Understand downstream licences, indemnity exclusions, permitted use and responsibility for prompts and outputs.

8. **Obtain jurisdiction-specific advice** before relying on fair use, a statutory exception or a disputed dataset for a material release.

Footnotes

1. *Bartz v. Anthropic*, Order on Fair Use, Dkt. 231, 23 June 2025 (court order reproduced by Justia).
2. *Bartz v. Anthropic*, Final Approval and Judgment, Dkt. 680, 20 July 2026; court-approved settlement documents.
3. *Kadrey v. Meta*, Partial Summary Judgment Order, Dkt. 598, 25 June 2025 (court order reproduced by Justia).
4. UK High Court, *Getty Images v. Stability AI*, [2025] EWHC 2863 (Ch), 4 November 2025.
5. Regulation (EU) 2024/1689, Article 53.
6. European Commission, GPAI Code of Practice and Copyright chapter.
7. California Legislature, AB 2013, chaptered text.
8. US Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability, January 2025.

11 / PRIVACY, DATA & SECURITY

Privacy, Data Governance, and Security

techne.ai/handbook/privacy-security/

Privacy and data governance are critical pillars of AI governance because AI systems consume and generate large volumes of data, often including personal and sensitive information. Compliance with privacy laws — GDPR in the EU, CCPA/CPRA and state-level laws in the US, PIPEDA in Canada, PIPA in Korea — is the starting point, but a mature AI governance programme also addresses data quality, model security, third-party risk, and incident response specifically calibrated to AI.

Foundational privacy law

The **EU General Data Protection Regulation (GDPR)** applies to in-scope processing of personal data, including training and inference. **Article 22** restricts solely automated decisions with legal or similarly significant effects, subject to exceptions and safeguards. It is not a general ban on automation. Related information duties appear in Articles 13–15. Establish the lawful basis, assess special-category data and complete a DPIA when required; adding a nominal human reviewer does not resolve every issue.[1]

In the United States, **CCPA/CPRA** provides California residents with rights concerning covered businesses, including access, correction, deletion, opting out of sale/sharing and limiting certain uses of sensitive personal information. Coverage, exceptions and the conditions on each right matter.[2] Other state privacy laws differ in thresholds, exemptions, automated-decision rules and effective dates. Maintain a jurisdiction-specific register rather than assuming California compliance covers every state.

Sector-specific regimes may also apply: HIPAA for covered entities and business associates, GLBA for covered financial institutions, FERPA for covered education records, and COPPA for covered online collection from children. Working in a sector does not automatically trigger every statute; map the entity, data and activity to the relevant rule. See Sectoral regulation.

International regimes include Brazil's LGPD, Japan's APPI, Singapore's PDPA and Korea's PIPA; their automated-decision rights are not interchangeable with GDPR Article 22. India's **DPDP Act and Rules have phased commencement following November 2025 notifications**: some provisions commenced then, with later tranches after 12 and 18 months. Do not describe the full substantive regime as already effective in 2024–2025.[3]

Data quality and lineage

Data quality influences model performance, but sound data alone does not ensure safe outcomes. A practical data-governance record includes:

• **Provenance tracking** — for each dataset, document source, acquisition method, licensing, and any consent obtained.

• **Quality assessment** — measure completeness, accuracy, freshness, representativeness; document known limitations.

• **Datasheets for datasets** — a documentation approach covering collection, composition, intended uses and limitations.[4] This can support evidence for applicable data-governance duties; EU AI Act Article 10 does not mandate the particular "Datasheets for Datasets" format.

• **Schema and version control** — treat training data with the same rigour as code.

Assess the scope and transitional dates of **California AB 2013** and **EU AI Act Article 53** before determining which public training-data documentation is required — see Copyright & IP.

Data minimisation and access control

Use data proportionate to the purpose and risk. Pseudonymisation, encryption and access controls can reduce exposure but do not automatically make personal data anonymous or establish a lawful basis. Consider:

• **Role-based access controls** restricting training-data access to authorised personnel and processes.

• **Data tokenisation** for training where individual records are not required.

• **Differential privacy** to bound privacy loss under a specified mechanism and privacy budget, with implementation and composition reviewed.[5]

• **Output-side restrictions** to reduce, not eliminate, the risk of revealing memorised personal data.

Privacy-enhancing technologies (PETs)

Privacy-enhancing technologies solve different problems. Select them against a documented threat model and measured performance constraints, not a blanket "privacy-preserving" label:

• **Differential privacy** — mechanisms such as DP-SGD or private aggregation; specify the protected unit, privacy parameters and cumulative privacy loss.[5]

• **Federated learning** — distributed training without centralising raw datasets. Model updates can still leak information; secure aggregation or differential privacy may be needed.

• **Homomorphic encryption** — computation on encrypted data; performance has improved but still constrains practical use to specific inference workloads.

• **Secure multi-party computation (MPC)** — computation designed to limit input disclosure under specified assumptions about the participants and protocol.

• **Trusted execution environments (TEEs)** — isolated execution with hardware and attestation dependencies; examine side-channel, administrator and supply-chain risks.

None of these techniques alone establishes anonymity, legal compliance or protection against every attack. For differential-privacy claims in particular, use NIST SP 800-226 to examine assumptions and implementation hazards.[5]

Retention and purpose limitation

Data governance policies should define **retention schedules** and **purpose-limitation controls** for training and inference data. GDPR requires data not be kept longer than necessary; CCPA permits consumer-initiated deletion. Practically:

• **Document retention windows** per dataset; automate deletion at end-of-window.

• **Re-purposing review** — if a dataset is reused for a new model or use case, evaluate consent and purpose limitation before training.

• **Deletion engineering** — identify affected corpora, retrieval stores, logs, backups and model artefacts. Evaluate applicable rights and exceptions, and document what can actually be removed. Machine unlearning is not a universally reliable substitute for retraining or a guarantee of legal erasure.

Model security

AI models themselves are attack targets. Threat categories include:

• **Model extraction** — adversaries query a model to approximate its behaviour or recover model information. Training-data extraction and membership inference are related but distinct privacy attacks.

• **Adversarial examples** — inputs crafted to cause misclassification.

• **Data poisoning** — tampering with training data to embed backdoors or biases.

• **Prompt injection** — manipulating LLM behaviour through crafted input, particularly via untrusted data in retrieval pipelines.

• **Model evasion** — bypassing safety filters or content controls.

The **NIST AI 600-1 Generative AI Profile** was published in **July 2024**. The separate **NIST AI 100-2 E2025**, published in **March 2025**, provides an adversarial-machine-learning attack and mitigation taxonomy. These are complementary voluntary resources, not a security certification.[67]

Defensive measures:

• **Adversarial training** — train on adversarial examples to harden the model.

• **Input and output filtering** — useful layers, not an authorisation boundary. Treat retrieved documents and tool results as untrusted data; enforce permissions in the application and tool layer.

• **Rate limiting and behavioural monitoring** — detect extraction and reconnaissance attempts.

• **Red-teaming** — scoped adversarial evaluation, with findings linked to mitigations and retests. Legal duties and voluntary commitments differ (see Frontier Models).

• **Provenance verification** of upstream components (base models, weights, fine-tuning datasets).

Data security

AI data pipelines must be secured with general infosec hygiene plus AI-specific considerations:

• **Encryption** in transit and at rest for training data and model weights.

• **Identity and access management** for systems handling training data and models.

• **Integrity verification** of training datasets (cryptographic hashing, version control).

• **Poisoning detection** — statistical anomaly detection in training data.

• **Backup and recovery** for model weights and training pipelines as critical assets.

Third-party and supply chain risks

Most AI systems incorporate third-party components: pre-trained foundation models, open-source libraries, cloud AI services, vector databases, datasets. Governance must extend to these:

• **Vendor assessment** for security, privacy, and AI-specific compliance posture.

• **Model provenance** — document source, version, and license of every model component.

• **License compliance** for open-source models with restrictive terms.

• **Contractual allocation** of responsibility for incident response, data handling, and regulatory cooperation.

• **Cross-border data transfer** mechanisms (Standard Contractual Clauses, adequacy decisions) where data crosses jurisdictions.

EU AI Act Article 25 explicitly addresses provider obligations through the value chain for high-risk systems.

Incident response

AI incident response should be a dedicated discipline within general incident response, with playbooks for:

• **Safety incidents** — AI causes physical or financial harm.

• **Bias incidents** — AI is found to produce discriminatory outcomes.

• **Privacy incidents** — AI reveals or is alleged to reveal personal data.

• **Security incidents** — AI is compromised or used to attack other systems.

• **Misuse incidents** — AI is used by adversaries for harmful purposes.

Map **each reporting regime separately**: the reporting entity, threshold, authority, clock start, deadline, initial-notice content and follow-up requirements. GDPR personal-data breach notification can involve a 72-hour supervisory-authority deadline, subject to its risk threshold. AI Act high-risk-system incident rules are not interchangeable with systemic-risk GPAI reporting, and California's critical-incident rules have their own timing. Do not use a generic "15-day AI incident" timer. See Frontier Models and Sectoral regulation.[1]

Independent reviews and red-teaming

Choose assessments that answer a defined question; distinguish legal requirements from procurement preferences:

• **Bias audits** — NYC LL 144 requires an independent audit for covered automated employment decision tools, not all hiring software.

• **Privacy and security reviews** — scope these to applicable law, contracts and the actual system; an annual review is not a universal statutory rule.

• **Adversarial evaluations** — selected frontier regimes require them; record method, coverage and residual risk.

• **ISO/IEC 42001 certification** — concerns a defined AI management-system scope, not a guarantee that each model is safe or compliant. ISO/IEC 42006 specifies requirements for certification bodies; it did not create the first possibility of certification. See ISO standards.

Coordination across functions

A practical programme coordinates **privacy, security, data governance, AI/ML engineering, legal, compliance and product** functions. A council or office is one organisational option, not a universally mandated structure. Assign decision authority, escalation and evidence ownership clearly; ISO/IEC 42001 offers a management-system approach where appropriate.

Footnotes

1. GDPR, Articles 5, 13–15, 22, 33 and 35.
2. California Attorney General, CCPA rights and scope.
3. MeitY, DPDP Rules 2025 and enforcement notifications; NIST, phased DPDP commencement overview.
4. Gebru et al., Datasheets for Datasets.
5. NIST, SP 800-226: Guidelines for Evaluating Differential Privacy Guarantees, March 2025.
6. NIST, AI RMF resources and AI 600-1 publication date.
7. NIST, AI 100-2 E2025: Adversarial Machine Learning, March 2025.

12 / TECHNICAL SAFETY

Technical Aspects of AI Safety

techne.ai/handbook/technical-safety/

AI governance needs technical evidence as well as policy. Engineering practices can reduce and reveal risk, but no checklist guarantees that an AI system is safe, reliable or aligned in every setting. This chapter offers a practical menu: choose controls for the system's capabilities, use context and potential harm, and record the limits of the evidence.

Robustness and reliability

Robustness begins with rigorous validation on test data that simulates real-world variability and edge cases. Practical techniques:

• **Stress testing** — computer vision systems tested in varying lighting and noise; text systems tested on out-of-distribution inputs and adversarial paraphrases.

• **Adversarial training** — training on adversarial examples to harden the model.

• **Ensemble methods** — multiple models or sensors may reduce some failures; correlated errors and shared dependencies can defeat the expected benefit.

• **Operating-domain documentation** — explicitly state what conditions the model was designed for; reject or flag inputs outside the domain.

• **Capability evaluations** — for frontier models, structured pre-deployment evaluations of dangerous capabilities (cyber, CBRN, autonomy). See Frontier Models for the structural pattern.

The **NIST AI 600-1 Generative AI Profile** dates to **July 2024**. Use it alongside the separate **March 2025 NIST AI 100-2 E2025** adversarial-machine-learning taxonomy; neither is a mandatory testing standard or an exhaustive account of current attacks.[1]

Alignment with human values

Alignment research and engineering seek to make AI behaviour better match intended objectives and constraints. Techniques include:

• **Reinforcement learning from human feedback (RLHF)** — widely deployed for instruction-following and value alignment in large language models.

• **Constitutional AI / RLAIF** — AI-assisted critique and revision against an explicit set of principles, reducing reliance on human raters at scale.

• **Direct preference optimisation (DPO)** and successors — offline preference learning that avoids the reinforcement-learning loop.

• **Specification through constraints** — encoding behavioural rules directly (e.g., refuse-list rules, scoped tool access).

• **Tool-use scoping** — for agentic systems, explicit permissions and human-in-the-loop checkpoints for high-impact actions.

For applications beyond instruction-tuned LLMs, alignment thinking still applies: define objective functions carefully (not just optimise the metric, but the metric subject to fairness and safety constraints), and peer-review model objectives during development.

For agentic AI systems — AI that takes actions through tools or other interfaces — add **controllability**: interruption, bounded permissions and recovery paths. Not every external action can be undone. The **International AI Safety Report 2026**, published **3 February 2026**, reviews capabilities, risks and the limitations of safeguards; it is scientific synthesis, not product approval.[2]

Interpretability and transparency tools

Interpretability sheds light on "black box" models:

• **SHAP and LIME** — per-prediction feature attribution; still standard for tabular and structured models.

• **Saliency maps and integrated gradients** — per-input attribution for vision models.

• **Attention visualisation** — can reveal attention patterns, but is not by itself a faithful causal explanation of an output.

• **Probing** — lightweight classifiers attached to model internals to test for specific learned representations.

• **Sparse autoencoders and circuit analysis** — research methods for examining internal features and computations. Human-readable interpretations can be incomplete or wrong and should be tested, not treated as a complete account of a model.

• **Model cards** — documentation in plain language describing intended use, performance, and limitations.[3]

Documentation requirements differ from the research convention of a **model card**. Relevant duties can include:

• EU AI Act Articles 11–13 for covered high-risk systems, subject to applicable dates (technical documentation, records and instructions).

• GDPR Articles 13–15 information duties and Article 22 safeguards for qualifying automated decisions.

• US sector-specific notice and explanation duties, including FCRA and ECOA where applicable.

A generic model card does not automatically satisfy any of these. See EU AI Act, Privacy and Sectoral regulation.

Bias and fairness mitigation

Measuring and mitigating bias is a discipline with multiple technical entry points:

• **Pre-processing** — examine representation, labels and sampling; reweight or resample when justified by the problem and evaluation evidence.

• **In-processing** — add fairness penalty terms or constraints to the objective.

• **Post-processing** — adjust model outputs to reduce disparity.

Evaluation libraries can help calculate metrics, but their defaults do not select a legally or ethically appropriate fairness definition. A useful **fairness review** records the population, protected groups where lawful to analyse, sample limitations, metrics, uncertainty, thresholds and remediation decisions.

Specific legal regimes now codify fairness expectations:

• **NYC Local Law 144** — bias audits for automated employment decision tools (see US State Laws).

• **EU AI Act Article 10** — data governance requirements for high-risk systems including measures to detect and prevent bias.

• **EU AI Act Article 27** — fundamental rights impact assessments for specified deployers and uses, not every high-risk deployment.

Engineering, legal, product and affected stakeholders should develop fairness criteria together. Metrics embed choices and can conflict under common conditions, such as different base rates; passing one numerical test is not proof that a system is lawful or free from discrimination. Record trade-offs and who approved them.[4]

Performance monitoring and drift detection

Deployment is not a one-time event. Governance requires ongoing monitoring:

• **Performance metrics** in production: accuracy, calibration, fairness metrics across subgroups, business metrics impacted by the AI.

• **Drift detection** — distribution shift in inputs (covariate drift), in labels (label drift), or in the input-output relationship (concept drift).

• **Out-of-distribution detection** — identify inputs unlike training data; route to human review or fall back gracefully.

• **Grounding and factual evaluation** for LLM applications — measure answer correctness and source support; retrieval can introduce inaccurate or malicious material and does not eliminate confabulation.

• **Misuse detection** — identify abuse patterns and adversarial attempts.

When monitoring signals a problem, governance triggers a **defined response process** (NIST RMF's "Manage" function): retrain, roll back, restrict, escalate. Document each incident and decision; feed findings back into design.

Safety constraints and testing in simulation

For AI interacting with the physical world (robots, autonomous vehicles, medical devices), pre-deployment safety testing is essential:

• **Simulation** — explore rare and dangerous scenarios while checking whether the simulation represents the operating environment. More simulated runs alone do not prove real-world safety.

• **Formal verification** — prove specified properties under explicit assumptions where tractable; a proof applies to its model and boundaries, not automatically to the complete deployed system.

• **Redundancy and fail-safes** — engineering patterns from aviation and industrial control extended to AI systems.

• **Guardian systems** — an independent monitor that can override or shut down the main AI if it detects unsafe behaviour.

For non-physical AI, apply **defence-in-depth**: pre-deployment evaluation, runtime controls, monitoring and incident response. The architecture in Frontier Models is one application of this pattern, not a universal safety benchmark.

Agentic systems: put controls around actions

For **agentic AI** — systems that take multi-step actions via tools, browsing, code execution or robotic control — govern what the system can do, not only what it can say:

• **Permission scoping** for tools and APIs the agent can call.

• **Human-in-the-loop checkpoints** for high-impact actions.

• **Action records** capturing relevant requests, tool calls, approvals and outcomes, with redaction and retention limits. Do not assume a model-generated rationale is a faithful account of its internal reasoning.

• **Sandboxing** of code execution and environment access.

• **Rate limiting** and **anomaly detection** at the action layer.

• **Reversibility analysis** — categorise actions by reversibility; require stronger controls for irreversible actions.

• **Identity and authorisation** — bind each action to a principal, scope and expiry; enforce permissions outside the model.

• **Untrusted-content separation** — documents, webpages and tool responses are data, not new authority to change the task or send secrets.

• **Transaction boundaries** — use idempotency, previews and explicit approval for consequential or irreversible actions; test partial failure and retries.

These are engineering recommendations, not a statement that every jurisdiction mandates the same agent controls. Test misuse and failure scenarios end to end, including permissions, downstream APIs, rollback and human escalation.

Verification, validation, and release evidence

A **safety argument** makes a bounded claim and exposes the evidence, assumptions and gaps supporting it. A safety case is one useful form; it is not identical to the transparency report required by California SB 53. Requirements and voluntary commitments must be checked separately. A practical structure includes:

1. The **claim** (e.g., “this model can be deployed for use case X with acceptable residual risk”).

2. The **evidence** — evaluations, independent reviews, monitoring data, third-party assessments.

3. The **argument** linking evidence to claim — explicit reasoning about how the evidence supports the claim.

4. The **counterarguments** — what could undermine the claim and how those risks are addressed.

Engineering and governance functions both contribute. Name the decision-maker accepting residual risk and the conditions that trigger reassessment. ISO/IEC 42001 can support this work as a voluntary management-system standard; adopting it or obtaining certification does not itself prove a particular system safe.

Footnotes

1. NIST, AI RMF resources and AI 100-2 E2025.
2. International AI Safety Report 2026, 3 February 2026.
3. Mitchell et al., Model Cards for Model Reporting, 2019.

4. Chouldechova, Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments, 2017.

13 / FRONTIER MODELS

Frontier Models

techne.ai/handbook/frontier-models/

Highly capable AI models attract additional governance attention, but **frontier model**, **general-purpose AI (GPAI)**, **systemic risk** and **high-impact AI** are not interchangeable legal categories. This chapter separates binding duties, voluntary compliance tools, testing partnerships and developer commitments. Classify the model and provider under each applicable regime before selecting the required evidence.

Figure: An illustrative map of governance mechanisms, not a hierarchy of legal equivalence. Research partnerships and voluntary frameworks do not replace statutory duties.

What counts as a “frontier model”?

There is no universal definition. Distinguish these criteria:

• **Compute** — EU AI Act Article 51(2) presumes high-impact capabilities above **10^{25} floating-point operations** of cumulative training compute; this is not a universal frontier-model definition. The Commission can amend the threshold by delegated act.

• **Capability** — ability to perform a wide range of distinct tasks, particularly tasks plausibly relevant to severe harm (CBRN, cybersecurity, autonomous replication, deception).

• **Designation** — the European Commission can designate GPAI models with systemic risk on capability or impact criteria; Article 52 also provides a process for a provider to substantiate an exceptional case against classification.[1]

A product name or a high benchmark score is not a legal classification. Record the exact model version, compute evidence where available, capabilities, distribution method, provider role and relevant designation. Reassess material updates; a static list of brand names becomes stale quickly.

EU GPAI obligations and the voluntary Code

Articles 53–55 establish GPAI duties, with scope-specific exceptions and transitional treatment. Depending on the model and provider, these include:

• Maintain **technical documentation** of the model (training, evaluation, capabilities, limitations).

• Provide **downstream documentation** to providers integrating the model into AI systems.

• Comply with **Union copyright law**, including TDM opt-outs.

• Publish a **summary of training content**.

• For **systemic-risk GPAI**: evaluations including adversarial testing, risk assessment and mitigation, serious-incident tracking/reporting, and cybersecurity.

The **GPAI Code of Practice**, published **10 July 2025**, is a voluntary route to demonstrating compliance; the underlying law is binding for in-scope providers. Its Safety and Security chapter concerns systemic-risk GPAI. Providers choosing another route still need adequate evidence. Check the Commission's current signatory list and chapter coverage: a signature is not independent proof that every model and practice complies. See EU AI Act for dates and exceptions.

CAISI testing agreements (United States)

On **5 May 2026**, NIST announced new CAISI agreements with **Google DeepMind, Microsoft and xAI**, building on earlier partnerships. These were not 2025 agreements. The announcement describes pre-deployment evaluation and research; CAISI's stated role includes voluntary agreements and unclassified national-security capability evaluations.[2]

These partnerships are **not a general US model-licensing requirement** or certification scheme. Do not infer unpublished contractual terms, mandatory release approval, or coverage of every model from a press announcement. For a participating provider, record the actual agreement, evaluation scope and follow-up obligations.

California SB 53 — Transparency in Frontier AI Act

California **SB 53**, signed **29 September 2025**, distinguishes a **frontier developer** from a **large frontier developer** (the latter includes an annual-revenue threshold). Its requirements include:[3]

• For large frontier developers, implement and publish a **frontier AI framework**, with specified risk and governance content.

• Publish a **transparency report** before or concurrently with a new or substantially modified frontier-model deployment, with additional assessment summaries for large developers.

• Report covered **critical safety incidents** to Cal OES within 15 days of discovery; imminent death or serious-injury risk has a separate 24-hour notification rule to an appropriate authority.

• Protect covered employee disclosures under its whistleblower provisions.

The statute does **not** simply require a document called a "safety case" for every release. A model card or other larger document can contain the required disclosures. Cross-mapping to an EU framework may save work, but does not establish equivalence. See US State Laws for scope.

Korea AI Basic Act — frontier safety track

Korea's AI Basic Act and Enforcement Decree took effect **22 January 2026**. MSIT describes a safety-obligation track requiring **all three** criteria: training compute above **10^{26} floating-point operations**, state-of-the-art technology, and a risk of broad and significant effects on fundamental rights. That track is **distinct from high-impact AI**, which depends on listed uses and risk. MSIT also announced an enforcement grace period with serious-harm exceptions; it is not a blanket exemption from every duty. Check the current Korean text, subordinate rules and territorial criteria rather than assuming every service accessible in Korea is identically covered.[4]

Voluntary frameworks and the AI Safety Institute network

Several voluntary mechanisms supplement statutory obligations:

• **Responsible Scaling Policies (RSPs)** — published by Anthropic and (in different forms) by other developers; commit to specific capability evaluations and risk thresholds.

• **Frontier Model Forum** — an industry organisation focused on frontier-AI safety research and shared practices; membership is not certification.

• **Safety benchmarks** — bounded evaluation datasets, not guarantees of safety outside the tested tasks. Record benchmark version, configuration, contamination risks and coverage gaps.

• **International AI Safety Report 2026**, published **3 February 2026** — an international scientific synthesis of GPAI capabilities, risks and safeguards, not binding regulation or a product endorsement.[5]

National evaluation organisations and international research partnerships can support shared methods and evidence. Their names, mandates and participation change; consult the responsible institution for the current programme. Do not treat network participation as a regulatory approval or assume that all members share every incident or evaluation result.

Common frontier-safety architecture

The following is this handbook's **practical synthesis**, not a claim that every cited regime mandates the same seven steps:

1. **Capability evaluations** at defined thresholds — before initial deployment, before scaling, after substantial updates.

2. **Risk identification and mitigation** — with documented thresholds at which mitigation actions trigger.

3. **Safety case** — structured argument that residual risk is acceptable, addressed to a defined audience (internal governance, regulator, public).

4. **Pre-deployment testing** — either internal or in cooperation with AI Safety Institutes.

5. **Post-deployment monitoring** — for misuse, incidents, capability change.

6. **Incident reporting** — map the actual recipient, threshold, deadline and confidentiality requirements for each applicable duty or agreement.

7. **Transparency** — published framework, safety case, model card.

Planning posture for frontier developers

Start with an applicability record and build proportionate evidence. Distinguish required actions from optional mechanisms:

• **EU GPAI compliance** where in scope, using the Code or another adequate route, with the relevant transition dates recorded.

• **Government testing partnerships** where available and appropriate; a CAISI agreement is not a universal condition for US operations.

• **California and Korean applicability reviews** using their separate model, developer and territorial criteria.

• **Published frontier AI framework / RSP** with specified capability thresholds and mitigation actions.

• **Release decision record**, potentially organised as a safety case, identifying evidence, residual risk and approval authority.

• **Management-system controls**, with ISO/IEC 42001 considered where useful or contractually required; certification is not a universal legal duty.

• **Incident reporting** procedures consistent with all applicable regimes.

For deployers and downstream integrators

If you integrate frontier models rather than develop them, the governance focus is:

1. **Evaluate upstream evidence**, not just a compliance badge: applicable disclosures, model limitations, testing scope, contractual rights and incident contacts.

2. **Receive and act on** downstream documentation provided under Article 53(1)(b) EU AI Act.

3. **Implement use-case-specific risk management** — the frontier-safety framework of the upstream developer does not substitute for your own risk assessment of the deployed application.

4. **Maintain provenance** — document which model version is in production, what changed when, and your validation of those changes.

5. **Monitor for incidents** in your deployment and feed back to the upstream developer.

Footnotes

1. Regulation (EU) 2024/1689, Articles 51–55.
2. NIST, 5 May 2026 agreement announcement and CAISI mandate.
3. California Legislature, SB 53, chaptered text, particularly Business and Professions Code sections 22757.11–22757.13.
4. MSIT, The AI Basic Act Comes into Force, January 2026.
5. International AI Safety Report 2026, 3 February 2026.

14 / GUIDANCE BY ROLE

Audience-Specific Guidance

techne.ai/handbook/audiences/

Different roles contribute different evidence and decisions. This chapter offers practical guidance for **AI practitioners**, **compliance officers**, **executives and board members**, and **policymakers and regulators**. These are suggested responsibilities to adapt to your organisation, not a universal allocation of legal liability.

AI Practitioners

Data scientists, ML engineers, AI developers

Focus

Practitioners are at the front lines of building and deploying AI. The job is to **operationalise** governance and safety inside the development process: build models that perform well on accuracy metrics *and* meet criteria for fairness, explainability, robustness, and compliance.

Specific practices

• **Translate principles into code.** Ethics guidelines mean nothing if they don't show up as concrete model-validation steps, bias checks, or model-card sections.

• **Handle data with a documented basis.** Establish applicable permissions and lawful bases, respect relevant rights and restrictions, and involve privacy reviewers early. Consent is not the only possible basis; pseudonymisation is not anonymity.

• **Test proportionately and record gaps.** Add stress tests, adversarial tests, fairness checks and, for relevant GenAI risks, red-teaming. No finite test suite exhausts all possible failures.

• **Maintain a system inventory.** Document purpose, owner, data, model version, dependencies, evaluations and deployment context. The inventory supports analysis; it is not proof of compliance.

• **Monitor in production.** Set up performance dashboards and drift alerts. Plan retraining cadence.

• **Partner with compliance early.** Check actual classification and dates for the EU AI Act, NYC LL 144 and relevant state rules. Colorado's SB 26-189 replaced the earlier SB 24-205 regime; do not design around the repealed framework. See US State Laws.

• **Cultivate a safety culture.** Ethical AI is everyone's job, like security. Speak up when something looks wrong; build feedback into development culture, not just review gates.

What changed for practitioners in 2025-2026

• **Frontier governance** combines different legal duties and voluntary commitments; a California transparency report is not identical to a safety case. See Frontier Models.

• **Training-data transparency** requirements have specific provider, release and jurisdictional scope. See Copyright & IP.

• **Agentic systems** add action-layer risks: permission scoping, reliable records, retry handling and reversibility analysis. The burden depends on capability and use, not the "agent" label alone.

Compliance Officers

Legal, regulatory, ethics, and risk personnel

Focus

Compliance officers ensure AI systems and processes adhere to law, regulation, and policy. They translate regulatory requirements into controls, guide AI projects, and verify controls are working.

Specific practices

• **Track the regulatory landscape.** EU AI Act (and Omnibus rebase), US federal EOs, state laws (CO, TX, CA, UT, IL, NYC), Korea AI Basic Act, Japan AI Promotion Act, China labelling rules, sector regulators. See Legal & Regulatory.

• **Develop internal policies.** AI governance policy, model-development standards, third-party AI usage policy, AI procurement standards.

• **Run role-specific training.** Teach classification, escalation and the actual controls each role must operate. Update training when legislation or deployment scope changes.

• **Review evidence.** Coordinate applicable data-protection and fundamental-rights impact assessments, independent bias reviews, model risk assessments and third-party contract review. Follow each requirement's actual scope.

• **Incident handling.** Coordinate response to compliance issues; regulator notifications (EU AI Office, Cal OES, state AGs); cross-functional incident triage.

• **Choose standards deliberately.** ISO/IEC 42001, 23894 and 42005 may support the programme. They are generally voluntary unless incorporated into an applicable obligation or contract; organisational size alone does not make them compulsory.

Key concerns

Map liability, regulatory exposure and operational harm to the actual actor and conduct. Penalty ceilings are not predictions of a likely fine, and statutes have different coverage, enforcement powers and cure provisions. Use the Legal & Regulatory chapters for source-backed details instead of a single headline penalty table.

What changed in 2025-2026

• **Federal preemption uncertainty** in the US complicates multi-state compliance — track the December 2025 EO and litigation outcomes.

• **Certification-body requirements** in ISO/IEC 42006 supplement management-system certification practice. Verify scope, accreditation and validity; certification is not regulatory approval.

• **Frontier-model rules** add a layer for large developers — see Frontier Models.

Executives & Board Members

C-suite, board directors, AI sponsors

Focus

Executives are responsible for **strategic oversight** and **organisational commitment** to AI governance. The job is to balance innovation with risk and to maintain stakeholder trust.

Specific practices

• **Set strategy.** Decide which AI use cases the organisation will pursue, which are out of bounds, and the corresponding risk appetite.

• **Establish governance structures.** AI Governance Council, model risk management function, AI ethics committee with real authority and budget.

• **Set the tone.** Communicate that responsible AI is a core value; reward responsible behaviour; back compliance teams when they say “not yet.”

• **Ask for decision-useful evidence.** Seek the material AI uses, responsible owners, unresolved risks, incidents, test limitations and overdue actions. Minutes should record what was considered and decided, not imply that a dashboard proves effective oversight.

• **Prepare for regulation.** Fund applicable requirements and track enacted-but-not-yet-applicable duties separately from proposals. Avoid treating every proposed “AI Bill” as law.

• **Evaluate credentials and partnerships separately.** ISO/IEC 42001 certification, GPAI Code signature and CAISI research agreements are three different mechanisms. Choose them for a reason and do not describe the latter two as certifications.

Considerations for 2025-2026

• **Public claims** — substantiate statements about responsible AI, environmental effects, fairness and safety; publish limitations alongside benefits.

• **AI workforce** — reskilling, AI literacy obligations under EU AI Act Article 4, internal AI usage policies.

• **Geopolitical risk** — export controls, data-localisation rules, jurisdictional fragmentation. The US December 2025 preemption EO and the ongoing state-federal tension affect operational planning.

What changed in 2025-2026

• **AI copyright** requires careful separation of acquisition, training and output risks. The *Bartz* settlement received final court approval in July 2026; it did not settle all AI copyright questions. See Copyright & IP.

• **Frontier developers** may face additional framework, transparency and whistleblower duties, depending on their statutory category. See Frontier Models.

Policymakers & Regulators

Government officials, regulators, standards body participants

Focus

Policymakers create and enforce the rules. The job is to address public risks while preserving innovation and accommodating sectoral diversity.

Specific focus areas

• **Develop and refine AI regulations.** EU AI Act implementation and Omnibus refinement, state legislative work, sector-specific rules.

• **Harmonise where possible.** OECD, G7 Hiroshima Process, ISO/IEC, IEEE; bilateral cooperation agreements (e.g., the International Network of AI Safety Institutes).

• **Build enforcement capacity.** Stand up AI offices and inspectorates; train staff; develop technical evaluation capability.

• **Address societal impacts.** Workforce displacement, AI literacy, public-sector AI use, election integrity, deepfake harms.

Concerns

Prevent harm; preserve fundamental rights; ensure national security; maintain transparency and accountability; balance with innovation; avoid over-regulation.

What changed in 2025-2026

• **Evidence sharing.** International scientific reports and testing partnerships can support common methods without creating identical legal duties. See Frontier Models.

• **Capacity building.** Invest in technical expertise, accessible complaint processes and methods for assessing real-world effects, not only published commitments.

• **Legal uncertainty.** Distinguish an executive direction, a filed lawsuit, an enacted amendment and a final court holding when communicating changes to the public. See US Federal and EU AI Act.

Cross-audience: the shared language

These audiences need a shared record, even when they use different frameworks:

• **An inventory and ownership map** identifying systems, purposes, people affected and accountable decision-makers.

• **An obligations register** separating law, contracts, internal policy and voluntary commitments.

• **A risk and evidence record** documenting testing, assumptions, unresolved gaps and residual-risk decisions.

• **An escalation and change log** showing how incidents, complaints and material releases prompt action.

NIST AI RMF and ISO management-system standards can help organise this work, but no single framework replaces applicable law. The handoff should tell the next team what is known, what remains untested and who decides — not simply attach a certificate or a maturity label.

15 / MATURITY PLANNING

AI Governance Maturity: An Illustrative Planning Framework

techne.ai/handbook/maturity-models/

This handbook offers six **illustrative, author-created planning frameworks**, each organised into seven stages. They are discussion aids for examining practices and choosing improvements, **not validated benchmarks, empirical maturity scores, certification criteria or legal conclusions**. NIST and ISO have not endorsed these stages. They cover:

1. **AI Governance Maturity** — organisational governance capability.
2. **AI Safety Maturity** — technical safety and reliability.
3. **AI Trust & Transparency Maturity** — stakeholder trust and transparency.
4. **Responsible AI Maturity** — ethics and social responsibility.
5. **AI Risk Management Maturity** — holistic risk management.
6. **AI Compliance Maturity** — processes for identifying and evidencing applicable obligations.

Use the descriptions to discuss the current state and plan improvements within a **defined scope**. Different systems and teams may fit different descriptions, and there is no proven time interval for moving between stages. Do not average the stage numbers, compare organisations without a common validated method, or infer that “Stage 7” means an AI system is safe or legally compliant. A low-risk use may need simpler controls than a high-impact use; more automation and more paperwork are not necessarily better governance.

Legal duties are minimum requirements when they apply, not achievements to postpone until a later stage. Certification, voluntary code signatures and public recognition are separate facts to verify; none is required to use this planning aid or sufficient to establish a stage. The stage headings below are shorthand for discussion, not findings about an organisation’s legal status or reputation.

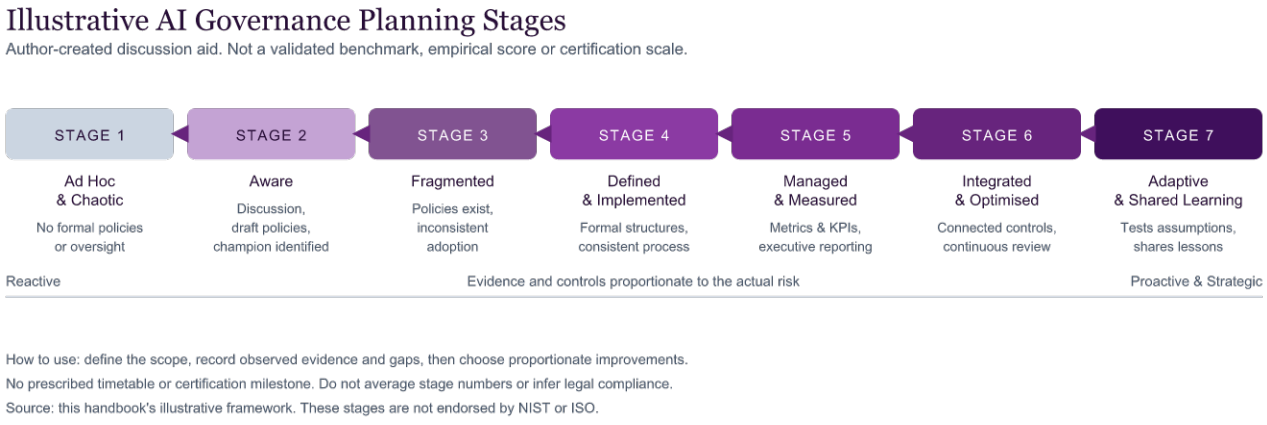

Figure: This handbook's illustrative progression. Stage labels organise discussion; they do not measure compliance, certify safety or prescribe a timetable.

1. AI Governance Maturity Stages

Stage 1: Ad Hoc & Chaotic

No formal AI governance. AI projects in silos with little oversight. Decisions on ethics or risk left to individual teams. No leadership awareness of AI-specific risk.

Assessment: No dedicated AI policies or roles exist.

Challenge: Lack of coordination — ethical or compliance breaches go unnoticed.

Best practice: Begin awareness building — basic AI risk workshop, inventory existing AI projects.

Stage 2: Aware (Initial Awareness & Planning)

The organisation has recognised the need for AI governance and is planning. Working groups form. Policies in draft. A champion may be advocating internally.

Assessment: Initial AI governance framework document; ethics committee formed (even without authority).

Challenge: Moving from talk to action.

Best practice: Roadmap with concrete milestones — publish AI ethics policy, assign roles, pilot procedures on one project.

Stage 3: Fragmented (Basic Policies, Inconsistent Adoption)

Basic policies exist; adoption is spotty. Some teams comply, others don't. Reviews for high-profile projects; many projects slip through.

Assessment: Policies on paper; some training delivered.

Challenge: Enforcement and coverage; viewed as box-ticking.

Best practice: Integrate governance into project lifecycle (sign-off gates); communicate success stories.

Stage 4: Defined & Implemented

Formal AI governance in place and functioning. Central committee or officer. Policies refined and communicated. Most projects follow required steps. AI governance is part of standard operating procedure.

Assessment: High percentage of AI initiatives follow the process; governance artefacts (risk assessments, model cards) exist per project. May target ISO/IEC 42001 alignment.

Challenge: Maintaining quality of execution; avoiding "compliance theatre."

Best practice: Internal reviews; investment in tooling that enforces process; named accountable executive.

Stage 5: Managed & Measured

Governance is measured and managed with metrics. KPIs are tracked (e.g., percentage of high-risk systems with completed FRIA, number of incidents, review findings closed). Process is refined based on data.

Assessment: Operational dashboards; regular reporting to executive leadership; tracked remediation pipelines.

Challenge: Avoiding metric overload; ensuring metrics reflect outcomes rather than activity.

Best practice: Tie incentives to governance outcomes; quarterly governance reviews with the board.

Stage 6: Integrated & Optimised

AI governance is integrated with quality, security, privacy and risk management. Changes in one function trigger reviews in the others. External certification may be relevant to a particular scope but is not a stage requirement.

Assessment: Evidence of cross-functional decisions, change management and completed corrective actions.

Challenge: Sustaining maturity through organisational change.

Best practice: Share findings externally; participate in standards bodies.

Stage 7: Adaptive Practice & Shared Learning

The organisation tests whether its governance remains effective as systems and risks change, and shares findings where appropriate.

Assessment: Independent challenge, documented adjustments after failures and evidence that findings influence later decisions.

Challenge: Avoiding complacency; staying ahead of evolving practice.

Best practice: Open-source governance tooling; publish a transparency report; mentor industry peers.

2. AI Safety Maturity Stages

Stage 1: Safety Practices Not Established

No deliberate safety practice. Models deployed without testing for adversarial inputs, drift, or failure modes.

Best practice: Establish minimum testing baseline; document known failure modes.

Stage 2: Reactive Safety Fixes

Safety issues addressed after they manifest. No proactive testing.

Best practice: Build incident response playbook; track recurring failure patterns.

Stage 3: Basic Testing & Validation

Standardised validation tests for new models. Some adversarial testing. Documented evaluation suites.

Best practice: Adopt NIST AI 600-1 GenAI Profile threat categories where applicable.

Stage 4: Proactive Risk Assessment & Mitigation

Risk assessment is standard for every AI project. Mitigations documented and tracked. Operating-domain documentation produced.

Best practice: Align with ISO/IEC 23894; produce model cards per system.

Stage 5: Advanced Technical Safeguards

Adversarial training, ensemble methods, guardian systems, formal verification of safety-critical components.

Best practice: Red-teaming as a standing practice; published evaluation results.

Stage 6: Continuous Safety Management

Continuous monitoring in production; drift detection; automated rollback. Safety incidents trigger root-cause analysis and feedback loops.

Best practice: Integrate safety telemetry into engineering dashboards; quarterly safety reviews.

Stage 7: Safety as a Differentiator

Safety arguments are challenged and updated as evidence changes. Evaluations include realistic failures and the limits of safeguards; absence of reported incidents is not proof of safety.

Best practice: Publish appropriately bounded safety findings and seek independent technical challenge where proportionate.

3. AI Trust & Transparency Maturity Stages

Stage 1: Limited Transparency

AI systems are black boxes; users have no information about how decisions are made.

Best practice: Begin publishing basic system descriptions; identify where transparency is legally required.

Stage 2: Basic Disclosures

Some disclosures — users informed they're interacting with AI; minimal information provided.

Best practice: Identify applicable disclosure and labelling duties, including EU AI Act Article 50 where in scope. Do not wait for a maturity stage to meet a legal deadline.

Stage 3: Explainability for Internal Use

Engineering teams use interpretability tooling (SHAP, LIME, saliency maps). Internal reviews of model decisions.

Best practice: Produce model cards; document operating domains.

Stage 4: User-Facing Explainability

End-users receive plain-language explanations of consequential AI decisions; appeals process available.

Best practice: Review applicable GDPR information and automated-decision safeguards; provide useful reasons and meaningful recourse where required.

Stage 5: Interactive Transparency & Engagement

Users can query AI decisions, provide feedback, and influence outcomes. Public transparency reports published.

Best practice: Consider content provenance measures and publish training-data summaries where required. Provenance does not establish that content is true.

Stage 6: Trusted AI Ecosystem

Transparency practices are informed by user feedback and tested for usefulness. Independent reviews may inform improvements; publication and certification do not automatically create trust.

Best practice: Engage with downstream stakeholders; publish responsible AI use cases.

Stage 7: Industry Transparency Leader

The organisation shares tested transparency practices and revises them in response to stakeholder challenge.

Best practice: Contribute to international standards (ISO, IEEE, C2PA); open-source tooling.

4. Responsible AI Maturity Stages

Stage 1: Responsibility Practices Not Established

No articulated ethics or responsibility principles. AI deployed without consideration of societal impact.

Best practice: Begin articulating principles; appoint ethics owner.

Stage 2: Articulated Principles (on Paper)

Ethics principles published; not yet operationalised. Risk of "ethics washing" without enforcement.

Best practice: Move from principles to processes; assign accountability.

Stage 3: Procedures and Training for Ethics

Ethics training rolled out; review procedures established; ethics escalation path defined.

Best practice: Make ethics review a default gate for AI projects.

Stage 4: Integrated Responsible AI Practices

Responsible AI practices integrated into product development. Bias mitigation, fairness checks, FRIA-style assessments standard.

Best practice: Conduct Article 27 FRIAs where the deployer and use are covered; use proportionate impact assessments elsewhere. ISO/IEC 42005 is a guidance resource, not a substitute for the applicable legal test.

Stage 5: External Accountability and Review

External reviews of ethics and responsibility. Independent ethics board with real authority. Public ethics commitments.

Best practice: Engage civil society; respond to external concerns.

Stage 6: Culture of Responsibility & Empowerment

Responsible AI is part of organisational culture. Employees feel empowered to raise concerns. Whistleblower protections in place (cf. California SB 53).

Best practice: Reward responsible decisions; protect whistleblowers; act on internal escalations.

Stage 7: Social Stewardship and Advocacy

The organisation actively advocates for responsible AI in the broader ecosystem. Funds research; supports public-interest initiatives (e.g., Current AI, ROOST).

Best practice: Sponsor open-source safety work; contribute to multilateral processes.

5. AI Risk Management Maturity Stages

Stage 1: No AI-specific Risk Management

AI risks not distinguished from general enterprise risks. No AI risk register.

Best practice: Create an AI risk register; inventory AI systems.

Stage 2: Qualitative Acknowledgment of AI Risks

AI risks identified at a high level. Documented but not quantified or mitigated systematically.

Best practice: Adopt NIST AI RMF as a starting framework.

Stage 3: Structured Risk Assessment Process

Standard process for risk assessment per AI project. NIST RMF Map and Measure functions implemented.

Best practice: Use NIST trustworthiness characteristics (privacy, accuracy, safety, fairness, etc.) as risk categories.

Stage 4: Risk Mitigation and Control Implementation

Risks have documented controls. NIST RMF Manage function implemented. Aligned with ISO/IEC 23894.

Best practice: Map controls to ISO/IEC 23894 risk treatment options.

Stage 5: Integrated Risk Management & Monitoring

AI risk integrated with enterprise risk management. Real-time monitoring of risk indicators. Cross-functional risk reviews.

Best practice: Quarterly AI risk reviews at executive level; aggregate dashboards.

Stage 6: Advanced Quantitative Risk Analysis

Quantitative risk models for AI — scenario analysis, sensitivity testing, financial risk modelling for AI-related losses.

Best practice: Choose quantitative methods only where data and assumptions support them; include uncertainty and qualitative scenarios rather than manufacturing precise loss estimates.

Stage 7: Adaptive and Resilient Risk Posture

Continuous improvement of risk practice. Resilience tested via tabletop exercises and red-team scenarios. Risk posture adapts to new threats (e.g., novel attacks against frontier models).

Best practice: Industry-leading incident-response drills; contribute to threat-intelligence sharing.

6. AI Compliance Maturity Stages

Stage 1: Obligations Not Mapped

The organisation lacks a documented view of applicable obligations. This label does not determine whether a legal violation has occurred.

Best practice: Review the current AI footprint against applicable regulations (EU AI Act, US state laws, sector rules).

Stage 2: Aware of Regulations

Aware of applicable rules but not yet implementing controls.

Best practice: Map regulations to AI systems; prioritise high-risk gaps.

Stage 3: Implementing Policies and Controls for Compliance

Policies and controls are being implemented against a defined obligations register. Evidence and unresolved gaps are tracked; implementation alone is not a compliance determination.

Best practice: Use ISO/IEC 42001 as architecture; close gaps systematically.

Stage 4: Comprehensive Compliance Management System

End-to-end compliance management. ISO/IEC 42001 aligned. Documented evidence per regulation.

Best practice: Integrate compliance evidence collection into engineering workflow.

Stage 5: Evidence Readiness and Independent Review

Evidence is organised for appropriately scoped external review. If certification is pursued, verify the certification body's accreditation and scope; ISO/IEC 42006 sets requirements for bodies certifying AI management systems, rather than itself accrediting them.

Best practice: Keep evidence current; choose independent review frequency according to obligations, risk and the applicable certification scheme.

Stage 6: Compliance as Business Enabler

Procurement and release decisions use a current, supportable account of obligations, controls, evidence and open gaps.

Best practice: Describe credentials accurately, including scope and validity; do not market certification or a code signature as regulatory approval.

Stage 7: Thought Leader and Shaper in AI Compliance

The organisation contributes experience to rule development and tests its own practices against new requirements and external challenge.

Best practice: Participate in standards development; contribute to regulator working groups.

How to use these frameworks

1. **Scope and evidence.** Name the systems, teams and date being considered. Record practices observed, documents examined, missing evidence and dissent; “not assessed” is a valid result.

2. **Prioritise.** Identify the framework where progress matters most for your organisation’s strategy and risk — often Compliance for regulated industries, Safety for frontier developers, Responsible AI for consumer-facing deployments.

3. **Plan.** Identify the specific practices needed to move to the next stage. Refer to relevant chapters: Legal & Regulatory, Technical Safety, Privacy, Data & Security, Frontier Models.

4. **Measure control effectiveness.** Track whether controls work, remediation is timely and affected people receive recourse. Counts of reviews or documents alone are activity measures; fewer reported incidents may reflect weaker detection.

5. **Reassess on change.** Review after material releases, incidents or legal changes and at a risk-appropriate interval. Do not assign a numerical score or promise an advancement timetable from these descriptions.

Relationship to established frameworks

The NIST AI RMF supplies a voluntary risk-management structure. ISO/IEC 42001 specifies an AI management system; ISO/IEC 42006 addresses certification bodies. Those resources informed the topics discussed here, **not the seven-stage scale**. Consult their actual requirements and current editions for implementation; the handbook does not reproduce or validate their full criteria.

16 / GLOSSARY

Glossary of AI Governance Terms

techne.ai/handbook/glossary/

These are **plain-language working definitions**, not verbatim ISO clauses or substitutes for statutory definitions. Similar words can have different meanings across frameworks. Follow the linked chapters and primary sources for the exact scope, conditions and applicable dates; classification depends on facts, not a glossary label.

Agentic AI

AI system designed to perform multi-step actions in pursuit of a goal, typically via tool use, code execution, or interaction with external systems. Distinguished from non-agentic AI by the ability to take autonomous actions beyond producing output for direct human use.

AI agent

A system that uses observations and available tools or actions to pursue objectives within an environment. Its actual autonomy depends on permissions, design and human oversight.

AI Act (EU)

Regulation (EU) 2024/1689 establishing harmonised AI rules. Entry into force, application dates and enforcement powers are distinct. See EU AI Act for the current timeline and amendment status.

AI component

A functional part of an AI system, such as a model, retrieval service, data-processing step or evaluation component. Its dependencies and role should be documented.

AI Office (EU)

A European Commission office supporting AI Act implementation, including GPAI oversight. It was established in 2024; 2 August 2025 was an application milestone for GPAI provisions, not its creation date. It does not replace every national competent authority. See EU AI Act.

AI Safety Institute / Center for AI Standards and Innovation (CAISI)

Organisations concerned with AI evaluation and safety or security research. The US CAISI and UK AI Security Institute have different mandates; a testing partnership is not automatically a regulatory approval. See Frontier Models.

Artificial intelligence (AI)

A broad field of engineered systems performing tasks associated with learning, inference, perception or reasoning. The precise legal definition depends on the instrument being applied.

Artificial intelligence system (AI system)

A deployed arrangement of models, data, software and other components producing outputs or actions. The system is not necessarily identical to its underlying model; applicable legal definitions may also address autonomy, adaptiveness and objectives.

Automated decision-making (ADM)

Decisions made or supported by automated processing, which need not use AI. “Solely automated” decision-making is a narrower legal category under GDPR Article 22, subject to qualifying effects, exceptions and safeguards. See Privacy.

Bias audit

An evaluation of specified disparities under a defined method. NYC Local Law 144 requires an independent bias audit for covered automated employment decision tools; it is not a complete finding that a hiring process is lawful or unbiased.

Conformity assessment

A process for demonstrating that specified requirements are fulfilled. EU AI Act routes for covered high-risk systems vary, including internal-control and notified-body procedures. ISO management-system certification is not automatically an AI Act conformity assessment.

Consequential decision

A term used in some laws for decisions affecting important opportunities or services. Its covered domains and thresholds depend on the statute. Do not apply the repealed Colorado SB 24-205 definition as current law; see US State Laws for the replacement regime.

Constitutional AI / RLAIF

Alignment technique using AI-generated feedback against an explicit set of principles to train models toward desired behaviour. Reduces reliance on human raters at scale.

Datasheet for dataset

Standardised documentation describing a dataset's motivation, composition, collection process, preprocessing, recommended uses, distribution, and maintenance. Proposed by Gebru et al. (2018); widely adopted as a transparency tool.

Differential privacy

A mathematical framework for bounding privacy loss under a specified mechanism, protected unit and privacy budget. Its guarantee depends on assumptions, implementation and composition; it is not a promise of zero disclosure. See Privacy.

Explainability

The ability to provide a useful, understandable account of a system's outputs or behaviour. An explanation can be plausible without being faithful; evaluate its audience, accuracy and limits.

Federated learning

Training across distributed datasets without pooling the raw records centrally. Shared updates can still reveal information, so privacy and security safeguards remain necessary.

Foundation model

A model trained on broad data that can be adapted for multiple tasks. “Foundation,” “general-purpose” and “frontier” overlap in everyday usage but are not interchangeable technical or legal classifications.

Frontier model

A context-dependent label for highly capable models. California SB 53 defines its own category; the EU uses GPAI and systemic-risk criteria. No single compute threshold defines “frontier” everywhere. See Frontier Models.

Fundamental Rights Impact Assessment (FRIA)

An assessment of effects on fundamental rights. EU AI Act Article 27 imposes a particular FRIA duty on specified deployers and uses. A general impact-assessment standard can help organise work but does not replace its legal requirements.

General-purpose AI (GPAI) model

The EU AI Act's model category based on significant generality and capability across distinct tasks, with further conditions and exclusions in Article 3(63). Not every GPAI model presents systemic risk. See EU AI Act.

GPAI Code of Practice

A voluntary tool to help demonstrate compliance with binding EU GPAI obligations. Its three chapters cover Transparency, Copyright, and Safety and Security; chapter coverage and implementation matter, not just signature status. See Frontier Models.

Hallucination

Generation of plausible but factually incorrect or fabricated output by a generative AI model. Mitigated through retrieval grounding, output verification, and clearer uncertainty signalling.

High-risk AI system (EU AI Act)

A system meeting Article 6 classification rules, including certain product-safety and Annex III uses. Merely appearing in a sector list does not answer every scope question; exceptions and application dates matter. See EU AI Act.

Impact assessment (AI)

Structured evaluation of an AI system's effects on individuals, groups, and society, covering risks, stakeholder analysis, fundamental-rights impacts, and mitigations. Standardised in ISO/IEC 42005:2025.

Interpretability

Understanding how a system or model produces its behaviour, using intrinsic structure or analytical methods. Its boundary with explainability varies by discipline; neither label guarantees a causal or complete explanation.

ISO/IEC 42001

A certifiable AI management-system standard published in 2023. Certification concerns a defined organisational scope, not proof that all AI products are safe or legally compliant. See ISO Standards.

Machine learning (ML)

Computational methods that fit patterns or decision rules from data or experience, rather than specifying every rule manually. Learning does not imply human-like understanding.

Model card

A documentation format describing intended use, evaluations, limitations and other model information. It can support required disclosures, but a model card is not automatically equivalent to technical documentation, a training-data summary or a statutory transparency report.

NIST AI Risk Management Framework (RMF)

A voluntary framework published in January 2023 with four functions: Govern, Map, Measure and Manage. The July 2024 Generative AI Profile (NIST AI 600-1) is a companion resource. Distinguish it from NIST cybersecurity publications and check each publication's draft or final status. See US Federal.

Reinforcement Learning from Human Feedback (RLHF)

Alignment technique training models using human preference signals. Widely deployed for instruction-following and value alignment in large language models.

Responsible Scaling Policy (RSP)

Frontier developer's published commitment to specific capability evaluations and risk thresholds, with mitigation actions triggered at defined thresholds. Originated with Anthropic; adopted in different forms by other frontier developers.

Risk-based regulation

An approach that calibrates duties to specified risks, actors or uses. Each law defines its own scope and thresholds; a low-risk label under one framework does not establish an exemption elsewhere.

Safety case

A structured claim, argument and supporting evidence about safety in a defined context, including assumptions and residual risks. It is not a guarantee of safety or identical to California SB 53's statutory transparency report. See Technical Safety.

Systemic risk (EU AI Act)

A defined EU GPAI risk category. Training above 10^{25} floating-point operations triggers a presumption of high-impact capabilities; Commission designation and the Article 52 procedure also matter. See Frontier Models.

Trustworthiness (in AI)

A context-dependent assessment of whether a system merits reliance, considering validity, reliability, safety, security, accountability, transparency, explainability, privacy and fairness. Trade-offs and evidence matter; no single score settles all of these.

Sources and scope

Use the chapter citations for legal terms. Foundational references include NIST AI RMF, Regulation (EU) 2024/1689 and ISO/IEC 42001. The Commission’s 2024 AI Office establishment decision and Colorado’s enacted SB 26-189 support the historical corrections above. These working definitions do not reproduce licensed ISO terminology or claim clause-by-clause conformity.

17 / RESOURCES

Resources

techne.ai/handbook/resources/

Start with the question, then choose the source

Use an official source to establish the rule, a dated explanation to understand its context, and a working record to document your decision. This directory is selective; inclusion is not an endorsement of a provider or a claim that a tool satisfies a legal requirement.

For detailed legal citations and qualifications, follow the source guide and the footnotes in the relevant chapter.

Official source collections

• **European Union:** European Commission AI Act policy hub for implementation material; EUR-Lex for the enacted text and amendments. Start with the handbook's EU AI Act chapter.

• **United States:** NIST AI Risk Management Framework, White House presidential actions and the relevant agency's official publications. The federal chapter distinguishes these different sources of authority.

• **State and local law:** use the legislature's enacted text, the current codified law and the responsible regulator's notices. The states and cities chapter links to the provisions discussed.

• **International comparisons:** the OECD AI Policy Observatory is a discovery aid. Verify legal status against each jurisdiction's official text; see international governance.

• **Standards:** the ISO artificial intelligence committee and official standard catalog entries identify editions and scope. Full standards may require a licensed copy; the ISO/IEC chapter does not reproduce their requirements.

Practical evaluation resources

These projects can support a defined evaluation, but no single metric resolves fairness, security or legal applicability.

• Fairlearn — tools and documentation for examining fairness in machine-learning systems.

• AI Fairness 360 — an open-source toolkit for examining and mitigating unwanted bias.

• Microsoft Responsible AI Toolbox — tools for evaluating aspects of model behavior.

• NIST AI Resource Center — supporting material for AI risk-management work.

Before adopting a tool, check its maintenance status, license, supported model and data types, privacy implications and fit to the decision. Record the version and limitations alongside the results. See technical safety for the evaluation workflow.

Continue within Techné AI

• **Understand a specific issue:** the public briefings connect selected legal sources to employment, board and governance questions.

• **Start an employment-AI inventory:** the free AI Hiring Exposure Snapshot helps organize initial questions; it is not a legal conclusion.

• **Use editable employment documents:** TalentSight Intelligence is a separate paid library. Review the current product page for its contents, price and terms.

• **Use editable board documents:** BoardSight Intelligence is a separate paid library, with its own product information and checkout.

• **Read the complete publication offline:** download the handbook PDF. It carries the same edition date as the web publication.

The handbook is free to read without an account. Buying a separate library is not required to access these chapters. See Terms for the website's terms and the separate paid-product policy.

Report an issue

Use the corrections process to report an outdated date, broken source, accessibility problem or substantive concern. Include the chapter URL, passage and a primary source where available. A report is not a promise of an individual legal or technical assessment.

18 / SOURCE GUIDE

Source Guide

techne.ai/handbook/references/

How to verify a statement

This is a reading map, not a claim that every source has been exhaustively reviewed or that all developments are covered. The **chapter footnotes are the detailed source record**: they identify the particular instrument, provision, publication or court order used for the statement.

When checking a legal proposition, record the instrument's status, relevant version, jurisdiction, regulated role, application date and any later amendment. Publication, entry into force, general application and a transitional deadline can be different dates. A press release about an agreement is not a substitute for the adopted text.

Law and regulation

Topic	Start with	Detailed handbook source record
European Union	Regulation (EU) 2024/1689 and Regulation (EU) 2026/1744, followed by relevant Commission implementation material	EU AI Act
US federal policy and agency requirements	Official executive orders, OMB memoranda, statutes and agency publications; distinguish internal federal requirements from private-sector duties	US federal
US states and cities	Current statutory text and official regulator publications, rather than only the bill announcement or an earlier effective-date article	US states and cities
Other jurisdictions	Official legislative and ministry sources; label pending bills and voluntary initiatives separately	International governance
Healthcare, finance and employment	The responsible sector regulator's statute, rule or dated guidance, including its stated scope and status	Sector-specific rules
Training data and copyright	The actual court order, procedural posture and affected claims; do not generalize one fact-specific ruling into a universal permission	Copyright and IP
Privacy and automated decisions	Applicable data-protection law and official regulator guidance, with attention to territorial scope and exceptions	Privacy, data and security

The September edition corrects the old handbook's use of secondary commentary where an official source was available. It also removes earlier forecasts that no longer describe the legal position. A source link is not a substitute for an organization-specific applicability review.

Standards and risk frameworks

• **NIST AI RMF 1.0:** official framework and supporting resources. See the federal and technical safety chapters for the distinction between voluntary risk guidance and binding obligations.

• **NIST AI 600-1:** Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, published July 2024. Do not confuse it with NIST's separate March 2025 adversarial-machine-learning publication.

• **ISO/IEC:** the standards chapter links the official catalog records for 42001, 23894, 22989, 42005 and 42006, and explains what can and cannot be established from public scope descriptions.

• **OECD:** Recommendation of the Council on Artificial Intelligence. Use the live official instrument to check its revision history and wording.

The handbook's maturity-planning tables are illustrative planning aids, not an official NIST scoring system or a validated assessment instrument. A claimed alignment should identify the specific source provisions and actual evidence, not just a framework name.

Technical evidence and practical examples

The technical safety, frontier-models and privacy chapters link technical and official materials beside the relevant discussion. Evaluate the study's task, data, system version, test conditions and limitations before applying a result elsewhere.

A vendor's description of its own policy is evidence of that published policy, not independent proof of the model's performance. A benchmark result is not a finding that every deployment context is acceptable.

Review boundaries

This edition is dated **September 7, 2026**. Sources were checked for the propositions discussed, but the handbook is not a continuously updated legal database, licensed standards collection, complete docket review or exhaustive global survey. Translations and summaries should be checked against the authoritative language where it controls.

For the changes made in this edition, see edition history. For an error or a later development, use the corrections process with the affected passage and source.

19 / EDITION HISTORY

Edition history

techne.ai/handbook/changelog/

v3.0.0 — September 7, 2026

The September edition brings the handbook into **techne.ai/handbook/** as part of the main website. It preserves the existing chapter structure and adds shared navigation, chapter contents, previous/next reading links, website breadcrumbs and connections to the relevant briefings, tools and libraries.

Source and copy corrections

• Updated the EU AI Act chapter for the adopted 2026 amendment and scoped transition dates, separating provider and deployer obligations.

• Replaced the superseded Colorado framework and added New York's amended RAISE provisions; corrected selected Texas, California, Utah, New York City and Illinois descriptions.

• Distinguished federal executive direction, agency obligations and statutes; corrected the NIST document/date conflation and current TAKE IT DOWN platform timing.

• Reviewed international and sector-specific descriptions against primary sources; removed unsupported predictions and assertions of comprehensive coverage.

• Corrected ISO/certification claims, copyright-case summaries, privacy/technical-safety descriptions and frontier-model governance distinctions.

• Identified the maturity frameworks as illustrative planning aids, not validated benchmarks or guaranteed paths to certification.

• Removed unsupported first/largest/outcome guarantees and stale model-release lists where they did not help the reader's decision.

Website and download

• The handbook now uses the main site's typography, colors, primary navigation, footer, Terms, Privacy and corrections links.

• Site-wide handbook links now point to the internal address instead of opening an isolated workers.dev publication.

• Chapter URLs remain individually addressable under /handbook/. Existing legacy chapter/anchor links are mapped to their equivalents.

• The complete PDF is regenerated from the same chapter content, with an edition date and source-parity check. No account, payment or Microsoft Forms submission is required to read or download this public reference.

The old host is retained for redirects after the new edition passes deployment checks; it is not the source of a separately maintained current edition. This publication remains distinct from paid library content and does not change member access rules.

v2.0.0 — May 11, 2026

The May edition restructured the original single-page handbook into a multi-page publication and added US state-law, copyright and frontier-model coverage. It was served separately on Cloudflare Workers. Some legal summaries, dates and certification statements in that edition were inaccurate or became superseded; the September edition above replaces them. Historical release descriptions are not current legal guidance.

v1.0.0 — March 2025

The initial public release introduced AI governance, ISO/IEC standards, EU AI Act concepts, NIST AI RMF, privacy, role-specific guidance, a glossary and six maturity-planning frameworks.

Maintaining an edition

Substantive corrections should update the affected chapter, its review date and this history. Regenerate and visually check the PDF when chapter content changes, and verify chapter links and old-URL redirects before publishing. A later date alone is not a source review.

Please use the corrections process to report errors with a precise passage and supporting source.

20 / ABOUT THIS PUBLICATION

About this publication

techne.ai/handbook/about/

The AI Governance Handbook is a public reference from **Khullani M. Abdullahi, J.D., founder of Techné AI**. It connects AI governance concepts with selected legal sources, technical safeguards and organizational decisions.

For the author's background and the firm's current work, visit About Techné AI. For the firm's service scope and disclosures, read Independence & Disclosures.

Purpose and review limits

The handbook is educational research, not a determination of a reader's legal obligations or a certification of an AI system. Qualified advisers should assess the specific facts, jurisdictions and sector when a consequential decision depends on the answer.

The September 2026 edition includes source-checked corrections and an integrated web/PDF release. AI tools assisted in the research, drafting and technical production. A chapter's review date describes that edition's source review; it does not promise continuous monitoring or imply independent professional sign-off. Full licensed ISO standards, every court docket and every jurisdiction have not been exhaustively reviewed.

Illustrative workflows and maturity stages are the publication's planning aids. They are not empirically validated benchmarks, official regulatory tests or guarantees of legal compliance, safety, certification or business outcomes.

Corrections and use

Use the corrections process to identify a passage, its URL and the source supporting a correction. Read the source guide and the footnotes in the relevant chapter before relying on a summary.

This publication is free to read and download. The site's Terms and Privacy Policy describe the website's general conditions. Paid TalentSight and BoardSight access is a separate offer with its own purchase and license terms.

Contact Techné AI for a defined research or advisory question.