AI LINE OF SIGHT

AIGP Foundations

Twelve modules across the AI Governance Professional Body of Knowledge



12 modules · 108 practice questions with explanations · about 9 hours of study

Legal content current as of November 2026

AI Line of Sight is an independent study resource. It is not affiliated with, endorsed by, sponsored by or approved by the International Association of Privacy Professionals (IAPP). IAPP and AIGP are trademarks of the IAPP, used here only to describe the certification this course helps you prepare for. This course is not legal advice and does not guarantee an exam result.
© 2026 AI Line of Sight · alltheseworlds.us

Contents

Each module: objectives, sections, exam tips, key terms, takeaways and a knowledge check with answers.

DOMAIN I
Module 1 AI Foundations: What AI Is and How It Works 3

DOMAIN I
Module 2 Why AI Needs Governance: Risks, Harms and Principles 13

DOMAIN I
Module 3 Roles and the AI Life Cycle 24

DOMAIN I
Module 4 Building an AI Governance Program 35

DOMAIN I
Module 5 Third-Party and Vendor AI Risk 46

DOMAIN II
Module 6 Privacy Law Applied to AI 57

DOMAIN II
Module 7 Other Existing Law: IP, Discrimination, Consumer Protection, Liability 69

DOMAIN II
Module 8 The EU AI Act and Other AI-Specific Laws 79

DOMAIN II
Module 9 Standards and Frameworks: OECD, NIST and ISO/IEC 91

DOMAIN III
Module 10 Governing AI Design, Data and Testing 101

DOMAIN III
Module 11 Release, Monitoring and AI Incidents 112

DOMAIN IV
Module 12 Governing AI Deployment and Use 122

Not affiliated with or endorsed by the IAPP. AIGP is a trademark of the IAPP.

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 1

AI Foundations: What AI Is and How It Works

Domain I of the AIGP Body of Knowledge · I.A.1 and I.A.3 · About 35 minutes

Every later topic in AI governance assumes you can tell an AI system from ordinary software, name the main kinds of AI, and follow how a model gets from data to a decision. This module builds that vocabulary. It is not a computer-science course: the goal is to understand AI well enough to ask the right governance questions about it.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Explain what makes a system an AI system, using the inference test that underpins the OECD and EU AI Act definitions.
2. Distinguish AI, machine learning, deep learning and generative AI, and place them in relation to each other.
3. Describe supervised, unsupervised, reinforcement and self-supervised learning, and give an example of each.
4. Outline how a machine-learning model is built, tested and used, and name what can go wrong at each step.
5. Use core terms precisely: model, training data, features, labels, inference, drift, hallucination, foundation model, agentic AI.

> **HOW TO USE THIS MODULE**
>
> Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. What counts as an AI system

There is no single legal definition of AI, but the definitions that matter most for governance have converged. The OECD revised its definition in 2023, and the EU AI Act (Article 3) adopted a closely aligned one. In plain terms, both describe an AI system as:

- a **machine-based system**,
- that works toward **explicit or implicit objectives**,
- and **infers**, from the input it receives, **how to generate outputs** such as predictions, content, recommendations or decisions,
- which can **influence physical or virtual environments**.

Both definitions add that AI systems vary in their **autonomy** (how much they act without human involvement) and **adaptiveness** (whether they keep changing after deployment).

The inference test

The word doing the work is **infers**. Traditional software executes rules a person wrote: if a claim is over USD 10,000, flag it. The logic came from a human. An AI system derives its logic from data or from a learned representation of the world: it was shown thousands of past claims and worked out for itself which patterns predict fraud.

Ordinary software	AI system
Follows rules written by people	Derives rules or patterns from data
Same input always gives the same output	Output can vary, and is often a probability
Behavior is fully specified in advance	Behavior emerges from training
Errors are usually bugs in the code	Errors can come from the data, the training or the context of use

EXAM TIP

Automation is not the test, and neither is making decisions. A spreadsheet macro that approves loans automatically is still not AI if a person wrote the rule. Watch for options that confuse "automated" with "AI".

2. AI, machine learning, deep learning and generative AI

These terms are nested, and exam questions often turn on using them precisely.

Term	What it means	Example
Artificial intelligence	The broad field of systems that perform tasks associated with human intelligence. Includes older rule-based expert systems as well as learning systems.	A chess engine, a recommendation system
Machine learning (ML)	A subset of AI in which systems learn patterns from data rather than being explicitly programmed.	A model predicting customer churn
Deep learning	A subset of ML using multi-layered neural networks. Powers most modern image, speech and language systems.	Image recognition, speech-to-text
Generative AI	Systems, usually built on deep learning, that produce new content: text, images, audio, code, video.	A chatbot drafting an email; an image generator

Narrow and general AI

Every AI system in use today is **narrow** (or weak) AI: it performs a defined task or range of tasks. **Artificial general intelligence** (AGI) refers to a hypothetical system matching human ability across virtually any intellectual task. AGI matters to governance mainly as a policy

debate; governance practice deals with narrow systems, including very capable general-purpose ones.

Discriminative and generative models

A **discriminative** model classifies or predicts: is this transaction fraud, what will this house sell for. A **generative** model produces new content resembling its training data. The governance concerns differ: discriminative models raise questions of accuracy and fairness in decisions about people, while generative models add questions about fabricated content, intellectual property and misuse.

3. How machines learn

Machine learning techniques are usually grouped by how the system learns. Each implies different data needs and different risks.

Technique	How it learns	Typical uses	Governance angle
Supervised learning	From labeled examples: inputs paired with the correct answer.	Credit scoring, spam filtering, medical image classification	Labels can carry human bias; quality of labels limits quality of the model.
Unsupervised learning	Finds structure in unlabeled data, such as clusters or anomalies.	Customer segmentation, fraud anomaly detection	Groupings can be hard to explain and may track protected traits.
Reinforcement learning	Learns by trial and error, receiving rewards or penalties for actions.	Robotics, game play, tuning chatbot behavior (RLHF)	The system optimizes the reward it is given, which may differ from what people actually want.
Self-supervised learning	Creates its own training signal from raw data, for example by predicting the next word.	Training large language and foundation models	Enormous, scraped datasets raise privacy, copyright and provenance questions.

RLHF (reinforcement learning from human feedback) is worth knowing by name: people rate a model's outputs, and those ratings are used to steer the model toward preferred behavior. It is a common step in making large language models helpful and safer.

4. From data to decision: the model life cycle

Whatever the technique, most ML systems pass through the same broad stages. Governance frameworks map controls onto these stages, so it pays to know them.

1. **Plan and design.** Define the problem, the intended use and users, and whether AI is the right tool at all.

2. **Collect and prepare data.** Source, clean and label data; select **features** (the input variables the model uses).
3. **Train.** The algorithm adjusts the model's internal **parameters** (or weights) to fit the training data.
4. **Validate and test.** Measure performance on data the model has not seen. Data is typically split into **training**, **validation** and **test** sets so the result is honest.
5. **Deploy.** Put the model into use, integrated with business processes and people.
6. **Operate and monitor.** Track performance, fairness and incidents; retrain or retire when needed.

Things that go wrong

Problem	What it means
Overfitting	The model memorizes its training data and performs poorly on new data. Strong test results on the training set prove little.
Underfitting	The model is too simple to capture the real pattern.
Data drift	The inputs the model sees in use shift away from what it was trained on (new customer groups, new products).
Concept drift	The relationship between inputs and outcome changes (fraud tactics evolve, so old signals stop working).
Data leakage	Information about the answer slips into training data, inflating test scores in a way that will not hold in real use.

> **WHY THIS MATTERS FOR GOVERNANCE**
>
> A model that passed testing is not finished. Drift means performance can degrade quietly after release, which is why monitoring and periodic review are core governance controls rather than optional extras.

5. Foundation models, generative AI and agents

Much current regulation and most current risk debate concern a newer layer of the technology.

Term	What it means
Foundation model	A large model trained on broad data at scale that can be adapted to many downstream tasks. Large language models (LLMs) are the best-known type.

Term	What it means
General-purpose AI (GPAI) model	The EU AI Act's term for a model that shows significant generality and can perform a wide range of distinct tasks. It carries its own set of obligations for providers.
Large language model (LLM)	A foundation model trained on text to predict the next token (a word or part of a word). Produces fluent text, code and analysis.
Multimodal model	A model that handles more than one type of data, such as text, images and audio together.
Fine-tuning	Further training a pre-trained model on a narrower dataset to specialize it.
Retrieval-augmented generation (RAG)	Supplying a model with relevant documents at the time of the question so its answer draws on that material, rather than on training data alone.
Prompt	The instruction or input given to a generative model. Prompt injection is an attack that hides instructions in content the model reads.
Hallucination	Fluent output that is false or unsupported, such as an invented citation. A product of how generative models predict likely text.
Agentic AI	A system that plans and takes actions toward a goal, using tools such as email, search or business software, with limited step-by-step human direction.

> **EXAM TIP**
> Agentic means acting, not just answering. A chatbot that looks things up and replies is not agentic; one that files the ticket, sends the email and updates the record is. Autonomy raises the governance stakes because errors become actions.

6. AI as a socio-technical system

Governance practitioners, and frameworks such as the NIST AI Risk Management Framework, treat AI as **socio-technical**: its behavior and its impact depend on the model, the data, the people who use it, the processes around it and the setting it is used in. The same model can be low-risk in one setting and high-risk in another.

Two consequences follow. First, many AI risks cannot be fixed purely with technical controls; they need policy, training, oversight and clear accountability. Second, governance has to involve more than data scientists: legal, privacy, security, risk, procurement and the business owners of each use all have a part. Module 2 turns to why those risks are distinctive.

Key terms

Term	Meaning
AI system	A machine-based system that infers from inputs how to generate outputs that can influence physical or virtual environments.
Model	The learned mathematical representation that turns inputs into outputs.
Training data	The data used to fit a model's parameters.
Feature	An input variable the model uses, such as age or transaction amount.
Label	The correct answer attached to a training example in supervised learning.
Inference	Using a trained model to produce an output from new input; also the defining capability in the OECD and EU definitions.
Drift	Degradation in performance because data or real-world relationships change after training.
Foundation model	A large model trained on broad data, adaptable to many tasks.
Hallucination	Generated output that is false or unsupported.
Agentic AI	AI that plans and takes actions toward goals using tools.

Key takeaways

- Inference, not automation, is what makes a system AI.
- AI contains ML, which contains deep learning; generative AI is a type of output, usually built on deep learning.
- The learning technique shapes the risk: labels carry bias, clusters resist explanation, rewards can be gamed, scraped data raises provenance questions.
- Models degrade after release through drift, so governance continues after deployment.
- AI is socio-technical: the same model's risk depends on who uses it, for what, and where.

Knowledge check

8 questions. Choose the single best answer for each.

1. A bank's software automatically declines any loan application where income is below a fixed threshold set by the credit team. Is it an AI system?

- **A.** Yes, because it makes decisions without a person
- **B.** No, because it applies a rule people wrote
- **C.** Yes, because it uses applicants' personal data
- **D.** No, because lending systems are excluded by law

2. Which statement best describes how these terms relate?

- **A.** Deep learning sits within ML, which sits within AI
- **B.** Machine learning is a subset of deep learning, not of AI
- **C.** Generative AI and AI are two names for one field
- **D.** AI is one branch of machine learning research

3. A streaming service groups viewers by viewing habits without any predefined categories. Which technique is this?

- **A.** Supervised learning
- **B.** Reinforcement learning
- **C.** Unsupervised learning
- **D.** Fine-tuning

4. A fraud model performed well at launch, but criminals changed tactics and its accuracy fell. What has happened?

- **A.** Overfitting to the training data
- **B.** Data leakage during testing
- **C.** Hallucination by the model
- **D.** Concept drift

5. Why is data split into training and test sets?

- **A.** To test performance on data the model hasn't seen
- **B.** To reduce the cost of storing and processing the data
- **C.** To satisfy data-minimization rules in privacy law
- **D.** To let two teams build competing models

6. A company connects its customer-service assistant to its ticketing and email systems so it can resolve issues on its own. What has it created?

- **A.** A foundation model
- **B.** An agentic system
- **C.** A multimodal model
- **D.** A discriminative model

7. A legal research assistant cites a court case that does not exist. What is this called?

- **A.** Data drift
- **B.** Prompt injection
- **C.** Hallucination
- **D.** Underfitting

8. Why do frameworks such as the NIST AI RMF describe AI as socio-technical?

- **A.** Because AI systems are built by social media firms
- **B.** Because AI models learn from social data alone
- **C.** Because regulators require human review of all outputs
- **D.** Because impact depends on people, process and context

Answers and explanations

#	Key	Why	Item ID · BoK
1	B	The logic was written by people, not inferred from data. Automation and personal data do not make a system AI, and no such lending exclusion exists.	FN-M1-01 · I.A.1
2	A	AI is the broad field; ML is one approach within it; deep learning is one family of ML techniques.	FN-M1-02 · I.A.1
3	C	Finding structure in unlabeled data is unsupervised learning. Supervised learning needs labels; reinforcement learning needs rewards.	FN-M1-03 · I.A.1
4	D	The relationship between inputs and the outcome changed after training. That is concept drift.	FN-M1-04 · I.A.3
5	A	Testing on unseen data gives an honest estimate of real-world performance and exposes overfitting.	FN-M1-05 · I.A.1
6	B	The assistant now plans and takes actions with tools toward a goal, which makes it agentic.	FN-M1-06 · I.A.1
7	C	Fluent but false output from a generative model is a hallucination.	FN-M1-07 · I.A.1
8	D	Risk and impact come from the interaction of the technology with the people, processes and setting around it.	FN-M1-08 · I.A.3

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 2

Why AI Needs Governance: Risks, Harms and Principles

Domain I of the AIGP Body of Knowledge · I.A.2 to I.A.4 · About 40 minutes

Organizations already govern software, data and vendors. This module explains why AI still needs governance of its own: what is distinctive about AI systems, the harms they can cause and to whom, where those harms come from, and the principles that responsible-AI frameworks use to guide decisions. It closes by showing how principles become practice, which is the subject of the rest of the course.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Identify the characteristics of AI that create distinctive risk: complexity, opacity, autonomy, speed and scale, probabilistic outputs, data dependency, adaptiveness and potential for misuse.
2. Classify harms as individual, group, societal or organizational, and recognize when one event produces several.
3. Name the main sources of AI risk, including types of bias, security threats, privacy and intellectual-property exposure.
4. Explain the core responsible-AI principles and where they come from (OECD, UNESCO, NIST).
5. Distinguish transparency, explainability and interpretability, and recognize when principles conflict.

HOW TO USE THIS MODULE

Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. What makes AI different

Most AI risks are versions of familiar risks: error, discrimination, privacy loss, security breach. What changes is how they arise and how far they spread. Eight characteristics account for most of the difference.

Characteristic	What it means	Why it raises risk
Complexity	Models with millions or billions of parameters, built from large pipelines of data and code.	Hard to test exhaustively; failures can come from unexpected interactions.
Opacity	It is difficult to see why a model produced a particular output (the "black box").	Errors and bias are hard to detect, explain or contest.
Autonomy	The system acts or decides without a person approving each step.	Mistakes take effect before anyone intervenes.
Speed and scale	Thousands or millions of decisions, made quickly.	One systematic error is repeated many times before it is noticed.
Probabilistic outputs	Outputs are likelihoods; the same input can produce different results.	Inconsistency, and confident-sounding errors.

Characteristic	What it means	Why it raises risk
Data dependency	Behavior reflects the data the model learned from.	Gaps and biases in data become gaps and biases in decisions.
Adaptiveness	Some systems keep learning or change after deployment.	Yesterday's test results may not describe today's system.
Potential for misuse	Capabilities can be used for purposes other than those intended (dual use).	Deepfakes, fraud, surveillance, or uses the developer never assessed.

EXAM TIP

Questions often describe a scenario and ask which characteristic is most at play. Look for the detail the scenario stresses: volume points to speed and scale, "no one can say why" points to opacity, "acted without sign-off" points to autonomy.

2. Who gets harmed: four levels of harm

A useful way to map AI harm is by who bears it. One event often produces harm at more than one level, and good answers name all that apply.

Level	Examples
Individual	Wrongful denial of a loan, job, benefit or medical treatment; privacy intrusion; physical injury from an unsafe system; loss of reputation from a deepfake.
Group	Discrimination against people who share a trait, such as race, sex, age, disability or dialect, even when that trait is not an input (because other features act as proxies).
Societal	Erosion of trust in information, misinformation and election interference, concentration of power, labor displacement, environmental cost of training and running large models.
Organizational	Legal and regulatory liability, fines, litigation, reputational damage, financial loss, loss of intellectual property, operational disruption.

The NIST AI Risk Management Framework uses a similar grouping: harm to **people** (individuals, groups and communities, society), harm to an **organization**, and harm to an **ecosystem** (interconnected systems, supply chains, the environment).

WATCH THE DISTINCTION

Harm to the organization is real, but it is a consequence for the company, not the harm the system did to the people affected. If a question asks what harm a biased hiring model causes, the answer is to candidates and groups first; the lawsuit is organizational harm on top.

3. Where AI risk comes from

Bias

Bias in AI is systematic error that disadvantages some people or groups. It can enter at every stage of the life cycle.

Type	How it arises
Historical bias	The data accurately reflects a world that was unfair, so the model learns past discrimination (for example, past hiring favoring one group).
Representation (sampling) bias	Some groups are under-represented in the data, so the model performs worse for them.
Measurement bias	The thing measured is a poor proxy for what matters (for example, using arrests as a proxy for crime, or healthcare spending as a proxy for need).
Aggregation bias	One model is applied to groups whose patterns genuinely differ.
Automation bias	People over-trust system output and stop exercising their own judgment.
Confirmation bias	Developers or users favor results that match what they already expected.

Security

AI systems inherit ordinary cybersecurity risks and add new ones:

- **Data poisoning:** corrupting training data so the model learns harmful behavior.
- **Adversarial examples (evasion):** small, crafted changes to input that cause a wrong output.
- **Prompt injection:** hidden instructions in content a generative model reads, overriding its intended behavior.
- **Model inversion and membership inference:** recovering training data, or learning whether a person's data was used, from the model's outputs.
- **Model extraction:** reconstructing a proprietary model by querying it repeatedly.

Privacy, intellectual property and reliability

- **Privacy:** large-scale collection, inference of sensitive traits from innocuous data, re-identification, and difficulty honoring rights such as deletion once data is in a model.
- **Intellectual property:** training on copyrighted material, outputs that reproduce protected works, uncertain ownership of AI-generated content, and leakage of trade secrets entered into tools.
- **Reliability and safety:** errors, hallucination, brittleness when conditions change, and failures in physical systems such as vehicles or medical devices.

4. Responsible-AI principles

Many organizations and governments have published AI principles. They differ in wording but converge on a common core. Three sources are worth knowing by name:

- **OECD AI Principles** (adopted 2019, updated 2024): the first intergovernmental standard on AI, endorsed by OECD members and others, and the basis for the G20 AI principles.
- **UNESCO Recommendation on the Ethics of AI** (2021): adopted by UNESCO's member states, with strong emphasis on human rights, dignity and sustainability.
- **NIST AI Risk Management Framework** (2023): a voluntary US framework whose characteristics of trustworthy AI are widely cited.

Principle	What it asks
Fairness	Avoid unjust discrimination; identify and manage harmful bias.
Safety and reliability	Perform as intended, consistently, without causing unacceptable harm, including under unexpected conditions.
Privacy and security	Protect personal data and resist attack and misuse.
Transparency	Be open that AI is being used, and about how and for what.
Explainability	Be able to give a meaningful account of why a system produced a particular output.
Accountability	Named people and organizations answer for the system and its outcomes.
Human-centricity and oversight	Serve human interests and rights, and keep meaningful human judgment where it matters.
Sustainability	Consider environmental and wider societal impact.

NIST frames trustworthy AI as **valid and reliable; safe; secure and resilient; accountable and transparent; explainable and interpretable; privacy-enhanced; and fair, with harmful bias managed**. Validity and reliability are treated as the base on which the others rest.

Transparency, explainability, interpretability

These three are often confused, and the difference is testable:

Term	The question it answers	Example
Transparency	Is it disclosed that AI is involved, and how the system works in general?	A notice that an AI system screened your application.
Explainability	Why did the system produce this outcome for this case?	The main factors that led to your application being declined.

Term	The question it answers	Example
Interpretability	Can people understand the model's internal mechanics?	A data scientist can read how each variable is weighted in a simple model.

> **EXAM TIP**
> Being told "AI assessed your claim" is transparency. Being told why it was declined is explainability. A scenario can meet one and fail the other.

5. When principles conflict

Principles are easy to endorse together and hard to satisfy together. Recognizing the trade-offs is part of the governance skill.

Tension	What it looks like
Accuracy and explainability	The most accurate model may be the hardest to explain.
Fairness and privacy	Testing for bias may require collecting sensitive data, such as ethnicity, that privacy principles would minimize.
Transparency and security	Revealing how a system works can help attackers or competitors game it.
Fairness definitions	Different mathematical definitions of fairness (equal outcomes, equal error rates) can be impossible to satisfy at once.
Human oversight and scale	Reviewing every decision may be impossible at volume, so oversight must be designed, not assumed.

The governance response is not to drop a principle quietly. It is to **name the trade-off, decide it at the right level of authority, and record the reasoning**, so the decision can be reviewed and defended.

6. From principles to practice

Principles alone change little. Governance is the set of structures and processes that turn them into decisions about real systems. The rest of this course builds that out, but the shape is worth seeing now:

- **Accountability:** clear roles, from board oversight to the owner of each AI use case.
- **Risk-based approach:** more scrutiny for higher-risk uses, rather than identical treatment of every tool.

- **Life-cycle coverage:** controls from design through monitoring and retirement, not only a pre-launch review.
- **Inventory and documentation:** knowing what AI the organization uses or builds, and recording key decisions.
- **Policies, training and culture:** people who understand the risks and know when to escalate.
- **Alignment with law and standards:** privacy and anti-discrimination law, AI-specific laws such as the EU AI Act, and frameworks such as the NIST AI RMF and ISO/IEC 42001.

Key terms

Term	Meaning
Opacity	Difficulty seeing why a model produced a particular output.
Proxy	A feature that correlates with a protected trait and can reintroduce bias, such as ZIP code for ethnicity.
Historical bias	Bias inherited from data that faithfully records an unfair past.
Automation bias	Over-reliance on system outputs by the people meant to check them.
Data poisoning	Corrupting training data to manipulate a model's behavior.
Prompt injection	Hidden instructions in content that override a generative model's intended behavior.
Transparency	Disclosure that AI is used and how it works in general.
Explainability	The ability to give reasons for a specific output.
Accountability	Named responsibility for a system and its outcomes.
Trustworthy AI (NIST)	Valid and reliable; safe; secure and resilient; accountable and transparent; explainable and interpretable; privacy-enhanced; fair.

Key takeaways

- AI's risks are mostly familiar risks made faster, larger, harder to see and harder to contest.
- Map harm by who bears it: individuals, groups, society, the organization. One event often hits several.
- Bias can enter at any stage, including through proxies and through the people who over-trust outputs.
- Responsible-AI principles converge across OECD, UNESCO and NIST; know their names and core content.
- Principles conflict. Good governance names the trade-off, decides it deliberately, and records why.

Knowledge check

8 questions. Choose the single best answer for each.

1. An insurer's model processes 50,000 claims a day. A single coding error affects every claim for a week before anyone notices. Which characteristic most magnified the harm?

- **A.** Speed and scale
- **B.** Opacity
- **C.** Adaptiveness
- **D.** Potential for misuse

2. A hiring model downgrades applicants from certain postcodes, which are home mostly to one ethnic group. Ethnicity is not an input. Which harm does the pattern add beyond harm to each applicant?

- **A.** Societal harm to the labor market
- **B.** Group harm through a proxy
- **C.** Organizational harm to the employer
- **D.** None, since ethnicity is not used

3. A health system used past healthcare spending as a stand-in for patients' medical need. Patients who historically had less access to care were scored as healthier than they were. Which bias is this?

- **A.** Automation bias
- **B.** Aggregation bias
- **C.** Measurement bias
- **D.** Confirmation bias

4. An attacker hides text in a web page that tells an AI assistant reading it to forward the user's emails. What is this attack called?

- **A.** Data poisoning
- **B.** Model extraction
- **C.** Membership inference
- **D.** Prompt injection

5. A bank tells a customer that an AI system assessed her loan, but cannot tell her why it was declined. Which statement is accurate?

- **A.** Transparency is met; explainability is not
- **B.** Explainability is met; transparency is not
- **C.** Both are met by the disclosure
- **D.** Interpretability is met; the others are not

6. Which intergovernmental AI principles were adopted in 2019 and later formed the basis of the G20 AI principles?

- **A.** The UNESCO Recommendation on the Ethics of AI
- **B.** The OECD AI Principles
- **C.** The NIST AI Risk Management Framework
- **D.** ISO/IEC 42001

7. To test a model for racial bias, the fairness team needs applicants' ethnicity, which the privacy team wants not to collect. What is the best governance response?

- **A.** Skip the bias test, since privacy law takes precedence
- **B.** Collect the data quietly without involving the privacy team
- **C.** Decide the trade-off at the right level and record why
- **D.** Drop the model entirely and revert to manual decisions

8. Claims handlers approve almost every recommendation an AI tool makes, even when they have information suggesting it is wrong. Which risk does this illustrate?

- **A.** Historical bias
- **B.** Data drift
- **C.** Model inversion attack
- **D.** Automation bias

Answers and explanations

#	Key	Why	Item ID · BoK
1	A	The error itself was ordinary; volume is what spread it to thousands of people before detection.	FN-M2-01 · I.A.3
2	B	Postcode acts as a proxy, so the harm falls on a group sharing a trait. The employer's legal exposure is a consequence, not the harm to applicants.	FN-M2-02 · I.A.2
3	C	Spending was a poor proxy for need, so the measure itself was biased.	FN-M2-03 · I.A.2
4	D	Hidden instructions in content the model reads are prompt injection. Poisoning targets training data.	FN-M2-04 · I.A.2
5	A	Disclosing AI use is transparency; giving reasons for this outcome is explainability, which is missing.	FN-M2-05 · I.A.4
6	B	The OECD AI Principles (2019, updated 2024). UNESCO's came in 2021; NIST's framework is a US voluntary framework; ISO/IEC 42001 is a management system standard.	FN-M2-06 · I.A.4
7	C	Principles conflict. The trade-off should be surfaced and decided deliberately, with reasons recorded, rather than one principle quietly abandoned.	FN-M2-07 · I.A.4
8	D	Over-trusting system output is automation bias. It undermines human oversight even when a human is formally in the loop.	FN-M2-08 · I.A.2

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 3

Roles and the AI Life Cycle

Domain I of the AIGP Body of Knowledge · I.B.5 and I.C.1 · About 40 minutes

The same AI system can create very different obligations for the company that built it, the company that sells it and the company that uses it. Many AIGP questions turn on this: before you can say what an organization must do, you need to know which role it plays and which stage of the life cycle the system is in. This module sets out the roles, shows how an organization can change role without meaning to, and maps governance controls onto each stage of the life cycle.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Distinguish developers, providers, deployers, users and affected persons, and name the EU AI Act's operator roles.
2. Recognize when an organization moves from one role to another, including by rebranding, substantially modifying or repurposing a system.
3. Describe the stages of the AI life cycle, including retirement, and the main governance control at each.
4. Explain how responsibilities divide between provider and deployer, and why a deployer cannot hand its accountability to a vendor.
5. Identify the events that should send a system back through governance review.

> **HOW TO USE THIS MODULE**
>
> Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. Why role comes first

Ask "what must this organization do?" and the answer usually starts with "it depends on its role". A bank that buys a credit-scoring model and the software company that builds and sells it both have obligations, but different ones. The builder answers for how the system was designed, trained, tested and documented. The bank answers for how it is used: on whom, with what oversight, and with what explanation to applicants.

Stage matters just as much. A duty that applies before release, such as testing against acceptance criteria, is not the answer to a question about a system that has been running for a year and started to drift.

> **EXAM TIP**
>
> Read every scenario for two signals before you look at the options: who is acting (builder, seller or user) and when (design, testing, deployment or operation). Many wrong options describe a real duty that belongs to a different role or a different stage.

2. The main roles

Role	Who it is	Example
Developer	Designs, builds or trains the AI system or model.	A data-science team training a fraud model; a lab training a foundation model.
Provider	Makes the system available to others, or puts it into use, under its own name. Often, but not necessarily, the developer.	A software company selling a hiring tool under its brand.
Deployer	Uses an AI system under its own authority in its own operations.	An employer using that hiring tool to screen applicants.
User or operator	The people who work with the system day to day inside the deployer.	Recruiters who review the tool's rankings.
Affected person	Someone the system's outputs are about, who may never touch it.	A job applicant scored by the tool.

The Body of Knowledge compares **developers, providers, deployers and users**. "User" is the slippery term: in everyday speech it can mean the deploying organization, its staff or a consumer using an AI product. Scenarios usually make the meaning clear from context.

The EU AI Act's operator roles

The EU AI Act defines its roles precisely and calls them, collectively, **operators** (Article 3). Simplified:

Role	Definition (simplified)
Provider	Develops an AI system or general-purpose AI model, or has one developed, and places it on the market or puts it into service under its own name or trademark, whether for payment or free of charge.
Deployer	Uses an AI system under its authority, except in the course of a personal, non-professional activity.
Importer	Established in the EU; places on the EU market a system bearing the name or trademark of a person established outside the EU.
Distributor	Any other party in the supply chain that makes a system available on the EU market without being its provider or importer.
Authorized representative	Established in the EU and mandated in writing by a non-EU provider to carry out obligations on its behalf.
Product manufacturer	Places an AI system on the market or puts it into service together with its own product, under its own name or trademark.

WATCH THE DISTINCTION

Under the EU AI Act the **provider**, not the developer, carries the main obligations. A consultancy that builds a model for a client, which then sells it under the client's brand, is the developer; the client is the provider. An organization that builds a system for its own internal use is both provider (it puts the system into service under its own name) and deployer.

3. How an organization changes role

Roles follow conduct, not contracts or intentions. Under Article 25 of the EU AI Act, a distributor, importer, deployer or other third party is treated as the **provider of a high-risk AI system**, with all of a provider's obligations, if it:

- **puts its name or trademark** on a high-risk system already on the market;
- makes a **substantial modification** to a high-risk system in a way that keeps it high-risk; or
- **changes the intended purpose** of a system that was not high-risk, including a general-purpose one, so that it becomes high-risk.

When that happens, the original provider is no longer treated as the provider of that specific system, but it must cooperate with the new provider and supply the information and access it needs.

Scenario	Role result
A retailer buys a general-purpose chatbot and uses it to answer product questions.	Deployer.
The same retailer starts using the chatbot to decide which job applicants get interviews.	Intended purpose changed into a high-risk (employment) use: treated as the provider of a high-risk system.
An HR software reseller sells a vendor's screening tool under its own brand.	Name or trademark added: treated as the provider.
A hospital retrains a vendor's diagnostic model on its own patient data, changing how it performs.	Possible substantial modification: may be treated as the provider. Needs careful assessment.

CURRENCY NOTE (AS OF NOVEMBER 2026)

Article 25 sits in the Act's high-risk chapter, whose application the Digital Omnibus on AI (Regulation (EU) 2026/1744) deferred: to 2 December 2027 for Annex III systems and 2 August 2028 for Annex I systems. Before the Omnibus those dates were 2 August 2026 and 2 August 2027, and the exam may still expect them. Module 8 covers the timeline.

4. The AI life cycle

Frameworks describe the life cycle in slightly different words. The NIST AI Risk Management Framework, the OECD and ISO/IEC 22989 cover the same ground: deciding whether to build, then data, model building, testing and deployment, then operation, monitoring and eventual retirement. A workable synthesis:

1. **Plan and design.** Define the problem, the intended purpose, the users and the people affected; decide whether AI is appropriate; classify the risk.
2. **Collect and process data.** Source, label and prepare data; establish provenance, rights, lawful basis and quality.
3. **Build and train.** Choose the approach, train and tune the model, and document design choices.
4. **Verify and validate.** Test for accuracy, robustness, fairness, security and fitness for the intended purpose. NIST calls this work **TEVV**: test, evaluation, verification and validation.
5. **Deploy.** Release into the operational setting, integrated with processes, people and human oversight.
6. **Operate and monitor.** Track performance, drift, incidents, complaints and use outside the intended purpose; maintain and update.
7. **Retire.** Decommission deliberately: tell users, decide what happens to data and models, keep the records the law requires.

> **WHY THIS MATTERS**
>
> The life cycle is a loop, not a line. Monitoring results, incidents and new uses send a system back to earlier stages. Governance happens at every turn, not once before launch.

5. Governance controls at each stage

Life-cycle policies (BoK I.C.1) attach a control, and the evidence that it happened, to each stage.

Stage	Main governance control	Typical evidence
Plan and design	Use-case intake, risk classification, impact assessment, decision to proceed	Intake record, risk tier, impact assessment
Collect and process data	Data governance: provenance, rights, privacy review, quality and representativeness checks	Data inventory, datasheet, privacy assessment
Build and train	Approved methods, documented design choices, reproducibility	Model documentation (a model card), experiment records

Stage	Main governance control	Typical evidence
Verify and validate	Testing against defined acceptance criteria, including fairness and security, by people independent of the builders where the risk warrants	Test reports, red-team results, sign-off
Deploy	Release approval gate, human-oversight design, user training, transparency notices	Approval record, instructions for use, notices
Operate and monitor	Performance and drift monitoring, incident handling, periodic review, change control	Monitoring reports, incident log, review minutes
Retire	Decommissioning plan, data retention or deletion, notice to users	Retirement record

Two ideas tie the stages together. **Gates** are points where named people must approve before a system moves on, with the level of authority scaled to the risk. **Accountability** means one named owner for each system, usually a business leader rather than a data scientist, who answers for it across its whole life.

Triggers for re-review

A system approved once can drift out of its approval. Policies should name the events that send it back through review:

- a new purpose, or a new population of affected people;
- retraining, a substantial model change, or a vendor's update to the underlying model;
- performance or fairness falling below agreed thresholds;
- a serious incident, a pattern of complaints, or a change in the law;
- the scheduled periodic review date.

6. Dividing responsibility between provider and deployer

Stage	Provider (builder or seller)	Deployer (using organization)
Plan and design	Defines the intended purpose and limits; assesses the risk of the design	Decides whether the use case is appropriate; assesses risk in its own context
Data	Quality, provenance and rights of the training data	Quality and relevance of the input data it supplies
Build and test	Designs, trains, tests and documents the system	Validates performance on its own population before relying on it

Stage	Provider (builder or seller)	Deployer (using organization)
Deploy	Supplies instructions for use and information about limitations	Follows the instructions; assigns trained human oversight; informs affected people where required
Operate and monitor	Monitors after release; fixes defects; reports serious incidents	Monitors its use; keeps the logs under its control; reports problems to the provider
Retire	Withdraws or recalls the system if necessary	Decommissions its use and handles its data

The deployer's column is the one most often missed. Buying a system moves the build work to the vendor, but not the accountability for how the system is used. A deployer that cannot explain a decision, did not validate a tool on its own population, or used it outside the vendor's instructions owns that failure. Module 5 covers vendor due diligence and contracts.

> **EXAM TIP**
> "The vendor is responsible" is a common distractor. Vendors answer for the design; deployers answer for the use. When a harm arises from how, where or on whom a bought system was used, look for the deployer's control.

Key terms

Term	Meaning
Provider (EU AI Act)	Develops a system or model, or has one developed, and places it on the market or puts it into service under its own name or trademark.
Deployer (EU AI Act)	Uses an AI system under its authority, outside personal, non-professional activity.
Importer	EU-established party placing a non-EU provider's system on the EU market.
Distributor	Any other supply-chain party making a system available on the EU market.
Operator	The EU AI Act's collective term for providers, deployers, importers, distributors, authorized representatives and product manufacturers.
Affected person	Someone a system's outputs are about, whether or not they use it.
Intended purpose	The use a provider designs a system for, as stated in its documentation and instructions for use.
Substantial modification	A change after release that the provider did not plan for and that affects compliance or the intended purpose.
TEVV	Test, evaluation, verification and validation: NIST's term for testing work across the life cycle.
Gate	A point in the life cycle where named people must approve before a system moves on.

Key takeaways

- Role first, then stage: most governance answers depend on both.
- Under the EU AI Act the provider, not the developer, carries the main obligations. An organization that builds for its own use is both provider and deployer.
- Rebranding, substantial modification or repurposing into a high-risk use can turn a deployer or distributor into a provider.
- The life cycle runs from planning to retirement and loops back. Each stage has its own control and evidence.
- Vendors answer for design; deployers answer for use. Accountability is not outsourced.

Knowledge check

8 questions. Choose the single best answer for each.

1. A consultancy builds a credit-scoring model under contract for a lender. The lender then sells access to it, under its own brand, to other banks. Under the EU AI Act, who is the provider?

- **A.** The consultancy, because it designed and trained the model
- **B.** The lender, which markets it under its own name
- **C.** The banks that buy access to the model
- **D.** No one until a conformity assessment is complete

2. A logistics company builds a route-planning model and uses it solely in its own operations. Which role or roles does it hold under the EU AI Act?

- **A.** Deployer, since it sells nothing to anyone
- **B.** Developer, since it does not market the system
- **C.** Both provider and deployer of the system
- **D.** Distributor, since it makes the system available to staff

3. A company deploys a vendor's general-purpose chatbot for customer queries, then starts using it to rank job applicants. What is the most significant consequence under the EU AI Act?

- **A.** It may be treated as the provider of a high-risk system
- **B.** It must stop using the chatbot for customer queries
- **C.** The vendor becomes the deployer of the hiring use
- **D.** It must register as a distributor of the vendor's chatbot product

4. A hospital uses an AI triage tool. Nurses review its scores, and patients are prioritized according to them. Who is the affected person?

- **A.** The nurse who reviews the score
- **B.** The vendor that supplied the tool
- **C.** The hospital, as deployer
- **D.** The patient being prioritized

5. Before any data is gathered, a governance office asks a team for the purpose of a proposed AI tool, the people it will affect and a risk tier. Which life-cycle stage is this?

- **A.** Plan and design
- **B.** Verify and validate
- **C.** Operate and monitor
- **D.** Collect and process data

6. Which event should send an approved AI system back through governance review?

- **A.** Routine patching of the server that hosts the model
- **B.** The vendor ships an update to the underlying model
- **C.** A new employee joins the team that operates it
- **D.** The annual budget for the system is approved

7. A bank bought a credit model and is sued after it is found to perform much worse for older applicants, a group the vendor did not test. Which control did the bank most clearly skip?

- **A.** Requiring the vendor to publish its source code
- **B.** Retiring the model when the contract ends
- **C.** Validating performance on its own applicants
- **D.** Registering the model in a public EU database

8. A company is replacing an AI claims model with a new one. Which action belongs to the retirement stage of the old model?

- **A.** Running a bias test on the replacement model
- **B.** Writing instructions for use for claims handlers
- **C.** Setting drift thresholds for the new model
- **D.** Deciding what happens to the old model's data

Answers and explanations

#	Key	Why	Item ID · BoK
1	B	The provider is whoever places the system on the market under its own name or trademark, here the lender. The consultancy is the developer; the banks are deployers.	FN-M3-01 · I.B.5
2	C	Putting a system into service under its own name for its own use makes it a provider; using it under its own authority makes it a deployer.	FN-M3-02 · I.B.5
3	A	Changing a system's intended purpose into a high-risk use such as employment can make the deployer the provider of a high-risk system (Article 25).	FN-M3-03 · I.B.5
4	D	The patient is the person the output is about. The nurse is a user, the hospital the deployer and the vendor the provider.	FN-M3-04 · I.B.5
5	A	Defining purpose, affected people and risk, and deciding whether to proceed, happens at plan and design, before data work begins.	FN-M3-05 · I.C.1
6	B	A change to the model itself can change performance and risk, so it is a re-review trigger. The other events do not change what the system does.	FN-M3-06 · I.C.1
7	C	A deployer should validate a bought system on its own population before relying on it. Vendor testing does not transfer the deployer's accountability for use.	FN-M3-07 · I.C.1
8	D	Retirement covers decommissioning the old system, including retention or deletion of its data and models. The other actions concern the new model.	FN-M3-08 · I.C.1

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 4

Building an AI Governance Program

Domain I of the AIGP Body of Knowledge · I.B.1 to I.B.4 and I.C.2 · About 45 minutes

Principles and role analysis matter only if someone in the organization is responsible for applying them. This module shows how to stand up an AI governance program: who does what, how functions work together, what an AI inventory and risk tiering are for, which policies to write or update, how to train people, and how to size all of it to the organization.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Describe the roles of the board, executives, an AI governance committee, system owners and supporting functions, and record them in a RACI.
2. Explain why AI governance must be cross-functional, and the mechanisms that make collaboration work.
3. Explain the purpose of an AI inventory and risk tiering, and what each should capture.
4. Identify the AI-specific policies to create and the existing policies to update.
5. Design role-based AI training, and describe the EU AI Act's AI-literacy provision.
6. Tailor a program to an organization's size, maturity, industry, risk tolerance and role.

> **HOW TO USE THIS MODULE**
>
> Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. What a governance program is

An AI governance program is the set of structures, roles, policies, processes and records through which an organization decides which AI to build or buy, under what conditions, and how it is overseen once in use. It is not a single committee or a single policy document.

Most organizations do not start from scratch. They extend what already works: the privacy program (impact assessments, records of processing), information security, enterprise risk management, procurement and, in banks, model risk management. Building on these makes AI governance faster to stand up and easier to run, provided the gaps AI creates are filled.

The three lines

Many organizations map governance onto the Institute of Internal Auditors' Three Lines Model:

Line	Who	Role in AI governance
First line	Business and technical teams that build, buy and use AI	Own and manage the risk of each system day to day
Second line	Risk, compliance, privacy, security, the AI governance office	Set policy, provide expertise, challenge and monitor
Third line	Internal audit	Independent assurance that the program works as designed

2. Roles and responsibilities

Role	Typical responsibilities
Board or board committee	Oversees AI strategy and risk; sets risk appetite; receives reporting on high-risk AI and serious incidents.
Executive sponsor	Senior owner of the program; secures resources; resolves escalations.
AI governance committee	Cross-functional body that sets policy, approves high-risk uses, and rules on escalations and exceptions.
AI governance lead or office	Runs the program day to day: intake, inventory, assessments, training, reporting.
System (use-case) owner	Business leader accountable for a specific system across its life cycle.
Data science and engineering	Build, test and document systems to policy; implement technical controls.
Privacy, security, legal, compliance	Specialist review: data protection, threats, legal exposure, regulatory duties.
Procurement	Applies AI due diligence and contract terms to vendors.
HR	Workforce impacts, AI in hiring and performance management, training.
Internal audit	Independent assurance over the program and high-risk systems.

The committee charter

A committee is only as useful as its charter. At minimum the charter should set out the committee's **mandate** (what it decides and what it only advises on), **membership** (by function, not just by name), **decision rights** (which risk tiers it must approve), **escalation paths** (what goes to the executive or the board), **quorum and meeting cadence**, and how decisions and dissent are **recorded**.

RACI

A RACI chart records who is **Responsible** (does the work), **Accountable** (owns the outcome; one person per activity), **Consulted** (gives input before the decision) and **Informed** (told afterwards). One row of a typical chart:

Activity	Responsible	Accountable	Consulted	Informed
Approve a high-risk AI use case	Governance office (prepares the file)	Committee chair	Privacy, security, legal, system owner	Executive sponsor, internal audit

> **EXAM TIP**
> Each system needs one accountable owner, usually a business leader. Options that leave accountability for a system with a group as a whole, or with the technical team that built it, are usually wrong.

3. Making it cross-functional

AI risk crosses disciplines. One hiring tool raises employment-law, privacy, security, fairness, procurement and reputational questions, and no single function can assess it alone. Mechanisms that make collaboration work:

- **One intake route.** A single form or portal for proposed AI uses, which routes each request to the right reviewers by risk tier.
- **Shared risk language.** A common taxonomy and set of risk tiers, so privacy, security and legal rate the same system in comparable terms.
- **Coordinated reviews.** Reviewers see each other's findings and contribute to one decision record, rather than signing off separately.
- **Clear escalation.** Disagreements and exceptions go to a named body with authority to decide.
- **Champions.** Trained contacts in business units who spot new AI uses early and route them in.

Failure pattern	What it looks like	Fix
Governance as a final hurdle	Reviews happen just before launch, when changing course is costly	Engage at intake and design
Silos	Privacy approves, security approves, nobody sees the combined risk	Coordinated review with one decision record
Tick-box compliance	Forms are completed but nothing changes	Attach conditions to approvals and check them in monitoring
Shadow AI	Teams adopt AI tools without telling anyone	Easy intake, acceptable-use rules, procurement and expense controls

4. Inventory and risk tiering

You cannot govern what you cannot see. An **AI inventory** (or register) records every AI system the organization builds, buys or uses, including AI features embedded in ordinary software. Typical fields:

- system name, purpose and accountable owner;

- role (provider, deployer or both) and vendor, if any;
- data used, including personal and sensitive data;
- people affected and decisions influenced;
- risk tier and the laws that apply;
- status, approval conditions and next review date.

Risk tiering sorts systems into levels, for example low, medium, high and prohibited, so that scrutiny matches risk. A low-risk drafting assistant might need registration and acceptable-use rules. A system that affects people's access to jobs, credit or care needs an impact assessment, committee approval and ongoing monitoring. Tiers should map to legal categories where those apply, such as the EU AI Act's, but an organization's own tiers can be stricter.

> **WHY THIS MATTERS**
> Proportionality keeps a program sustainable. Treating every tool as high-risk overwhelms reviewers and pushes teams to work around governance; treating everything as low-risk misses the systems that can do real harm.

5. Policies: what to write and what to update

Most organizations need a few new AI policies and a larger number of updates to existing ones (BoK I.C.2).

New

- **AI policy or standard:** scope, principles, roles, risk tiers, the approval process and the consequences of not complying.
- **Acceptable use of AI tools:** approved tools, prohibited inputs (confidential, personal or client data), checking outputs, disclosure of AI use.
- **Procedures** for intake, impact assessment, testing, approval, monitoring, incidents and retirement.

Updated

Existing policy	What AI changes
Privacy	AI processing as a trigger for impact assessments; lawful basis for training data; automated decision-making; honoring data-subject rights for data in models.
Information security	AI-specific threats (prompt injection, data poisoning, model theft); access to models and training data; security of AI vendors.
Data governance	Provenance and quality of training data; retention of training and input data; records of the data used.

Existing policy	What AI changes
Intellectual property	Confidential information in prompts; ownership of AI-generated output; licensed or copyrighted material in training data.
Procurement and third-party risk	AI due-diligence questions and contract clauses (Module 5).
Records and HR	Retention of AI decision records; AI in hiring and performance management.

> **EXAM TIP**
> When a question asks for the most efficient response to a new AI risk, updating the existing policy that already covers the area usually beats writing a parallel standalone one.

6. Training and awareness

Policies work only if people know them. Effective AI training is **role-based**:

Audience	Focus
All staff	Acceptable use, recognizing AI risks, protecting data, how to report a concern.
Users and operators of specific systems	The system's limitations, automation bias, their oversight duties, when to override or escalate.
Developers and data scientists	Responsible-AI methods, documentation, bias and robustness testing, the approval process.
Procurement, legal, privacy, security	AI-specific due diligence and review techniques.
Leaders and the board	AI strategy and risk, risk appetite, the questions to ask, their oversight duties.

Track completion, refresh training when policies, tools or laws change, and test understanding rather than attendance.

AI literacy under the EU AI Act

Article 4 of the EU AI Act places an AI-literacy duty on **providers and deployers**, covering their staff and others who operate or use AI systems on their behalf, taking into account their knowledge and experience and the context in which the systems are used.

CURRENCY NOTE (AS OF NOVEMBER 2026)

As adopted, Article 4 required providers and deployers to take measures to **ensure**, to their best extent, a sufficient level of AI literacy, from 2 February 2025. The Digital Omnibus on AI (Regulation (EU) 2026/1744) softened this from 27 July 2026 to a duty to take measures to **support** AI literacy, without guaranteeing any particular level for any individual. The exam may expect the original wording.

7. Tailoring the program

There is no single correct program. Governance should fit the organization (BoK I.B.4):

Factor	How it shapes the program
Size and resources	A small firm may fold AI review into an existing privacy or risk committee; a large group may need a dedicated office.
Maturity	Early stage: inventory, acceptable use and intake first. Mature: full life-cycle controls, metrics and possibly ISO/IEC 42001 certification.
Industry and regulation	Regulated sectors such as finance, health, insurance and employment need tighter controls and more documentation.
Risk tolerance (appetite)	Sets which uses are off-limits, which need senior approval and how much residual risk is acceptable.
Role	A developer or provider needs strong build-and-test controls; a deployer of bought tools needs strong procurement, validation and use controls.
Geography	The jurisdictions in scope decide which laws the program must meet.

Operating models

Model	How it works	Trade-off
Centralized	One central team reviews and approves all AI	Consistent, but can become a bottleneck far from the business
Federated	Business units govern their own AI to common standards	Close to the business, but consistency varies
Hub and spoke (hybrid)	A central hub sets policy and handles high-risk cases; units handle the rest	Balances both; needs clear rules on what goes to the hub

Whatever the model, measure it: inventory coverage, the share of high-risk systems with completed assessments, time from intake to decision, training completion, and incidents and how they were resolved. Metrics show the board whether governance is working and where to invest.

Key terms

Term	Meaning
AI governance program	The structures, roles, policies, processes and records through which an organization decides on and oversees AI.
AI governance committee	Cross-functional body that sets AI policy and approves high-risk uses.
System owner	The business leader accountable for one AI system across its life cycle.
Charter	The document setting a committee's mandate, membership, decision rights, escalation and records.
RACI	Responsible, Accountable, Consulted, Informed: a chart of who does what for each activity.
Three Lines Model	IIA model separating risk ownership (first line), oversight (second) and independent assurance (third).
AI inventory	A register of every AI system the organization builds, buys or uses.
Risk tiering	Sorting systems into risk levels so that scrutiny matches risk.
Acceptable use policy	Rules on how staff may use AI tools, including prohibited inputs.
Hub and spoke	Hybrid operating model: a central hub sets policy and handles high-risk cases.

Key takeaways

- Build on existing privacy, security, risk and procurement programs, and fill the gaps AI creates.
- Every system needs one accountable owner; a committee charter sets decision rights and escalation.
- AI risk is cross-functional: one intake route, shared risk language, coordinated reviews.
- An inventory and risk tiers make governance proportionate.
- Update existing policies first; add an AI policy and acceptable-use rules.
- Train by role, and tailor the program to size, maturity, sector, risk appetite, role and geography.

Knowledge check

8 questions. Choose the single best answer for each.

1. A bank's new AI governance committee has agreed its membership but keeps deadlocking over who can approve a high-risk use. What is it most likely missing?

- **A.** A larger budget and more support staff
- **B.** A charter setting decision rights and escalation
- **C.** More data scientists among its members
- **D.** An external audit of every pending use case before approval

2. Who should be accountable for an AI claims-triage system across its life cycle?

- **A.** The data-science team that trained it
- **B.** The AI governance committee as a body
- **C.** The vendor that supplied the base model
- **D.** The head of claims, its business owner

3. Privacy, security and legal each approved a hiring tool separately. None saw that, together, their concerns made it high-risk. What would best prevent this?

- **A.** Coordinated review with one decision record
- **B.** Asking each function to review more slowly
- **C.** Moving all review to the data-science team
- **D.** Requiring board sign-off on every AI tool purchased

4. Claims handlers who use an AI tool keep accepting its recommendations without question. Which training would help most directly?

- **A.** General awareness training for all staff on AI concepts
- **B.** Training on the tool's limits and when to override
- **C.** Technical training on how the model was built
- **D.** Board briefings on the firm's AI risk appetite

5. Which parties does the EU AI Act's AI-literacy provision (Article 4) address?

- **A.** Providers and deployers, for their own staff
- **B.** National regulators, for their inspectors
- **C.** Affected persons, before using an AI service
- **D.** Importers and distributors, for their customers

6. A 40-person firm uses three off-the-shelf AI tools and has no AI governance yet. What is the most sensible first step?

- **A.** Seek ISO/IEC 42001 certification for the firm
- **B.** Hire a dedicated AI governance team of five
- **C.** List its AI tools and set rules for use
- **D.** Ban all AI tools until a full program exists

7. Staff are pasting client contracts into a public generative-AI tool. Which policy action addresses this most directly?

- **A.** A new model-validation standard for data science
- **B.** A retention schedule for model training data
- **C.** Acceptable-use rules on prohibited inputs
- **D.** Add the tool to the risk-tier taxonomy

8. A company's privacy policy predates its use of AI. It now plans to train models on customer data. What is the most efficient governance response?

- **A.** Write a standalone AI privacy policy alongside the old one
- **B.** Pause all AI work until a new privacy program exists
- **C.** Rely on the vendor's privacy policy instead
- **D.** Update the privacy policy to cover AI training

Answers and explanations

#	Key	Why	Item ID · BoK
1	B	Membership alone does not say who decides. A charter sets decision rights and escalation paths, which is what breaks a deadlock.	FN-M4-01 · I.B.1
2	D	Each system needs one accountable owner, usually the business leader whose process it serves. Teams and committees cannot hold single-point accountability.	FN-M4-02 · I.B.1
3	A	Siloed sign-offs miss combined risk. Coordinated review feeding one decision record lets each function see the whole picture.	FN-M4-03 · I.B.2
4	B	This is automation bias among operators. Role-based training on the tool's limitations and when to override addresses it directly.	FN-M4-04 · I.B.3
5	A	Article 4 addresses providers and deployers, covering their staff and others operating AI systems on their behalf. The Digital Omnibus softened the duty's wording but not who it applies to.	FN-M4-05 · I.B.3
6	C	A small, early-stage deployer should start proportionately: know what it uses and set acceptable-use rules. Certification and a dedicated team come with maturity and scale.	FN-M4-06 · I.B.4
7	C	An acceptable-use policy that prohibits entering confidential or client data into unapproved tools targets the behavior itself.	FN-M4-07 · I.C.2
8	D	The existing privacy policy already governs customer data. Updating it to cover AI training avoids parallel, conflicting rules.	FN-M4-08 · I.C.2

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 5

Third-Party and Vendor AI Risk

Domain I of the AIGP Body of Knowledge · I.C.3 · About 40 minutes

Most organizations buy far more AI than they build. AI now arrives inside software they already use, through features switched on by an update, and through tools employees sign up for themselves. Buying a system moves the build work to a vendor, but not the accountability for how it is used. This module covers the policies, assessments, contract terms and acceptable-use rules that let an organization govern AI it does not control.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Explain why third-party AI creates risks beyond those of ordinary software procurement, including the layered AI supply chain.
2. Describe a third-party AI life cycle from identifying the need to exit, and the policy that governs it.
3. Identify the main areas of AI vendor due diligence, and prefer independent evidence over vendor assertion.
4. Name the contract terms that matter most for AI, and what to do when bargaining power is limited.
5. Write acceptable-use rules for third-party AI tools, and check the organization's own use against the vendor's terms.
6. Explain how to monitor a vendor's AI after contract, including the triggers for re-assessment.

HOW TO USE THIS MODULE

Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. Why third-party AI is different

Traditional vendor risk management asks whether a supplier is secure, solvent and compliant. Those questions still matter for AI, but AI adds problems that ordinary procurement does not catch:

Problem	What it means for the buyer
Opacity	The buyer usually cannot inspect the model, its training data or how it was tested.
Changing behavior	The vendor can update or replace the underlying model, so the system the buyer approved may not be the system it is running.
Data reuse	Inputs, prompts and outputs may be retained or used to train the vendor's models unless the contract says otherwise.
Embedded AI	AI features appear inside existing products, sometimes enabled by default in an update, without a new purchase that would trigger review.
Shadow AI	Staff sign up for AI tools with a credit card or a free account, outside procurement altogether.

Problem	What it means for the buyer
Layered supply chain	One vendor's product is often built on another company's foundation model, hosted by a third, so risks sit several layers away from the buyer.
Concentration	Many vendors rely on the same few model providers, so one provider's failure or policy change affects many systems at once.

The AI supply chain

A typical chain has three or more links: a **foundation model provider** trains a general-purpose model; an **application vendor** builds a product on it, often with fine-tuning and its own data; the **deploying organization** buys the product and uses it. The vendor's subcontractors are your **fourth parties**. Due diligence has to reach far enough down the chain to know which model is used, where data goes and who can change things.

> **WATCH THE DISTINCTION**
> Outsourcing the build does not outsource the accountability. A deployer that relies on a vendor's assurances without validating the system on its own population, or uses it outside the vendor's instructions, owns the resulting harm (Module 3).

2. The third-party AI life cycle and policy

Third-party AI is governed through a life cycle that sits alongside procurement rather than inside a single purchase decision:

1. **Identify and classify.** Record the proposed use in the AI inventory and assign a risk tier. Any product with an AI feature counts, not just products marketed as AI.
2. **Due diligence.** Assess the vendor and the system to a depth that matches the tier.
3. **Contract.** Negotiate terms that allocate data, IP, performance, change and incident responsibilities.
4. **Onboard and validate.** Configure the system, test it on the organization's own data and population, train users, set up human oversight.
5. **Monitor.** Track performance, incidents, vendor changes and the vendor's own risk posture.
6. **Exit.** Retrieve or delete data, transition to a replacement, and keep the records needed.

A **third-party AI policy** (or an AI section in the existing third-party risk policy) makes this repeatable. It should say what triggers review (including AI features added to existing products), who reviews at each tier, minimum due-diligence and contract requirements, who may approve exceptions, and how often each tier is re-assessed.

EXAM TIP

When a scenario describes an AI feature quietly switched on in software the organization already licenses, the governance gap is usually the trigger: the policy did not treat a new AI feature as a change that needs review.

3. Due diligence

AI due diligence extends the usual security and privacy questionnaire with questions about the model itself. The depth should match the risk tier.

Area	Questions to ask	Better evidence
Purpose and limits	What is the system designed for? What uses does the vendor rule out? What are its known limitations?	Instructions for use, model or system card
Data	What data was the model trained on? Does the vendor have the rights to it? Will our data be used for training?	Data provenance statement, contract terms
Performance and fairness	How was it tested, on what populations, with what results? Has it been tested on people like ours?	Test reports, independent bias audit
Security	How is the system protected, including against AI-specific attacks such as prompt injection and data poisoning?	SOC 2 Type II report, ISO/IEC 27001 certificate, penetration-test summary
Privacy	Is the vendor a processor or a controller for our data? Where is data stored and transferred? Who are the sub-processors?	Data processing agreement, sub-processor list, transfer mechanism
Governance	Does the vendor run an AI management system? How does it handle incidents and model changes?	ISO/IEC 42001 certificate, change and incident procedures
Regulatory status	Is the system high-risk under the EU AI Act or similar laws? Can the vendor supply what a deployer needs to comply?	Conformity documentation, registration details
Supply chain and viability	Which foundation model does it use? Which fourth parties are involved? Is the vendor financially stable?	Architecture summary, fourth-party list, financials

Evidence over assertion

A vendor's "yes" on a questionnaire is a claim, not evidence. Prefer, in rough order: independent certifications and audits, test results you can reproduce or that a third party produced,

documentation, and finally self-attestation. Then **validate the system yourself** on your own data before relying on it: no vendor test covers your population and context exactly.

> **EXAM TIP**
> A SOC 2 Type II report or ISO/IEC 27001 certificate shows that security controls exist and operate. It says little about accuracy, bias or fitness for your use. Options that treat a security attestation as proof a model is fair or reliable are wrong.

4. Contracts

The contract is where due-diligence findings become enforceable commitments. Clauses that matter most for AI:

Clause	What it should cover
Data use	No use of the customer's inputs, outputs or data to train or improve the vendor's models without explicit agreement; retention and deletion periods; confidentiality.
Privacy roles	Controller and processor roles, a data processing agreement, sub-processor approval, transfer mechanisms.
Intellectual property	Ownership of outputs; warranties about training-data rights; indemnity against third-party IP claims.
Performance	Agreed performance measures, service levels, and remedies when they are missed.
Transparency and documentation	Instructions for use, documentation of limitations, and information the customer needs to meet its own legal duties.
Change notification	Advance notice of model replacements, material updates and new AI features, with the right to test or refuse them.
Incidents	Notice of security breaches and AI incidents within set times; cooperation in investigation.
Audit and cooperation	Rights to audit or receive audit reports; cooperation with regulators and with the customer's assessments.
Flow-down	Equivalent obligations imposed on the vendor's subcontractors and model providers.
Liability and exit	Liability caps that reflect the risk; termination rights; data return or deletion; transition assistance.

When you cannot negotiate

Large model providers often offer standard terms only. A buyer with little bargaining power still has options: choose the enterprise tier whose terms exclude training on customer data, read the published terms and acceptable-use policy closely, restrict which data and uses the tool is approved for, add internal controls such as output review, and record the residual risk and who accepted it.

> **CURRENCY NOTE (AS OF NOVEMBER 2026)**
> Under the EU AI Act, a provider of a high-risk system and the third parties that supply it with AI tools, services or components must agree in writing on the information, capabilities, technical access and assistance the provider needs (Article 25(4)). That article sits in the high-risk chapter, which the Digital Omnibus deferred to 2 December 2027 for Annex III systems and 2 August 2028 for Annex I systems (originally 2 August 2026 and 2027).

5. Acceptable use of third-party AI

Most employees meet third-party AI through general-purpose tools: chat assistants, writing aids, coding assistants, meeting transcribers. An **acceptable-use policy** for these tools typically covers:

- **Approved tools and accounts:** which tools are allowed, and that work happens in enterprise accounts under the organization's contract, not personal or free accounts.
- **Prohibited inputs:** confidential information, client data, personal data and sensitive data, unless the tool is approved for them.
- **Prohibited uses:** decisions about people (hiring, credit, discipline) without going through governance review; uses the law restricts.
- **Checking outputs:** people remain responsible for accuracy, citations and code they use.
- **Disclosure and IP:** when to disclose AI use, and how to treat AI-generated material whose ownership or originality is uncertain.
- **Reporting:** how to report a problem or a new tool someone wants to use.

The vendor's acceptable-use policy binds you too

Model providers publish their own usage policies that prohibit certain uses, such as some high-stakes automated decisions or use in specified sectors without safeguards. A deployer must check that its intended use complies; breaching them can mean suspension of service as well as legal exposure.

6. Ongoing monitoring and exit

Approval at contract signature is a snapshot. Ongoing oversight of a vendor's AI includes:

- monitoring the system's performance and fairness on the organization's own cases, not the vendor's benchmark;
- tracking vendor notices of model changes, and re-testing before or soon after they take effect;
- collecting updated audit reports and certificates when they renew;
- periodic re-assessment at a frequency set by the risk tier.

Events that should trigger an early re-assessment: a model replacement or major update, new AI features, a serious incident or breach at the vendor, a change of ownership or a key fourth party, a change in the law, or a change in how the organization uses the system.

Plan the **exit** before signing. Lock-in is a real risk with AI: prompts, configurations, fine-tuned models and workflows may not transfer. Know how data will be returned or deleted, how long a transition would take and what the fallback is if the service stops.

Key terms

Term	Meaning
Third party	An external organization that supplies a product or service, here one with AI in it.
Fourth party	A vendor's own subcontractor or supplier, such as the foundation model provider behind an application.
Embedded AI	AI features built into software the organization already uses.
Shadow AI	AI tools adopted by staff without approval or review.
Due diligence	Assessment of a vendor and its system before contracting, scaled to risk.
SOC 2 Type II	An independent report on whether a service provider's security controls operated effectively over a period.
Data processing agreement	The contract that binds a processor to handle personal data only on the controller's instructions.
Change notification clause	A contract term requiring the vendor to give notice of model changes and new AI features.
Acceptable-use policy	Rules on how staff may use AI tools, including approved tools and prohibited inputs.
Lock-in	Dependence on a vendor that makes switching costly or slow.

Key takeaways

- Third-party AI adds opacity, changing behavior, data reuse and a layered supply chain to ordinary vendor risk.
- Treat a new AI feature in existing software as a change that triggers review.
- Prefer independent evidence over vendor assertion, and validate the system on your own population.
- Contracts should settle data use, IP, performance, change notice, incidents, audit and exit.
- Acceptable-use rules govern staff use of AI tools, and the vendor's own usage policy binds the deployer too.
- Monitor after signing; a vendor model change is a re-assessment trigger.

Knowledge check

8 questions. Choose the single best answer for each.

1. A CRM vendor's routine update switches on an AI feature that drafts customer emails. Nobody at the customer reviewed it. What is the most likely governance gap?

- **A.** The vendor had no current SOC 2 Type II report
- **B.** AI features added by updates were not a review trigger
- **C.** Staff had not been trained in how to write customer emails
- **D.** The CRM contract lacked a cap on vendor liability

2. A vendor's questionnaire says its hiring model is "fully tested for bias". What should the buyer do next?

- **A.** Seek independent evidence and test on its own applicants
- **B.** Accept the answer, since the vendor bears legal liability
- **C.** Ask the vendor to repeat the claim in writing
- **D.** Rely on the vendor's ISO/IEC 27001 certificate instead

3. A company is about to sign with a generative-AI vendor and wants to stop its confidential inputs being used to improve the vendor's models. Which clause matters most?

- **A.** An indemnity against third-party IP infringement claims
- **B.** A service-level agreement on uptime
- **C.** A data-use clause barring model training
- **D.** A cap on the vendor's liability

4. Which evidence best shows that a vendor's AI system is secure, rather than accurate or fair?

- **A.** The vendor's published model card
- **B.** A bias audit on a similar population
- **C.** Its benchmark accuracy scores
- **D.** A SOC 2 Type II report

5. A small firm wants to use a major model provider's chatbot but cannot negotiate its standard terms. Which is the most reasonable response?

- **A.** Use an enterprise tier and restrict data and uses
- **B.** Abandon AI use until it can negotiate custom terms
- **C.** Use free personal accounts to avoid the contract
- **D.** Accept the terms without any review, since they are standard

6. An HR team's hiring tool is built on a large language model from a different company. What is that model provider to the HR team's organization?

- **A.** Its controller for applicant data
- **B.** A fourth party in its supply chain
- **C.** Its distributor under the EU AI Act
- **D.** Its deployer of the hiring tool

7. A vendor tells customers it will replace the model behind its fraud-scoring product next month. What should a customer do?

- **A.** Nothing, since the contract is unchanged
- **B.** Terminate the contract immediately
- **C.** Re-test the system on its own cases
- **D.** Ask for a lower price for the new model

8. A marketing team plans to use a provider's image generator to create ads. Before approving the use, what should governance check besides internal policy?

- **A.** The provider's headcount in the EU
- **B.** The provider's credit rating
- **C.** The team's past campaign results
- **D.** The provider's own usage policy

Answers and explanations

#	Key	Why	Item ID · BoK
1	B	Embedded AI often arrives by update. The third-party policy should treat a new AI feature as a change requiring review.	FN-M5-01 · I.C.3
2	A	Self-attestation is the weakest evidence. Independent audit results and the buyer's own validation on its population are stronger, and the deployer stays accountable for use.	FN-M5-02 · I.C.3
3	C	Restrictions on using customer data to train or improve the vendor's models belong in the data-use clause. The other clauses address different risks.	FN-M5-03 · I.C.3
4	D	A SOC 2 Type II report evidences operating security controls. It does not show accuracy or fairness, which need test results and audits.	FN-M5-04 · I.C.3
5	A	Limited bargaining power is compensated by choosing protective terms where offered, restricting use, adding internal controls and recording who accepted the remaining risk.	FN-M5-05 · I.C.3
6	B	A vendor's own supplier is a fourth party. Due diligence should reach far enough down the chain to know which model is used and where data goes.	FN-M5-06 · I.C.3
7	C	A model replacement can change performance and risk, so it is a re-assessment trigger. The customer should test on its own cases before or soon after the change.	FN-M5-07 · I.C.3
8	D	Model providers' usage policies restrict some uses. The deployer must confirm its intended use complies, or risk suspension and legal exposure.	FN-M5-08 · I.C.3

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 6

Privacy Law Applied to AI

Domain II of the AIGP Body of Knowledge · II.A.1 to II.A.4 · About 50 minutes

Privacy law is usually the first law an AI system meets. It applies whenever personal data is processed, at every stage: when training data is collected, when the model is built, when inputs are fed in and when outputs make decisions about people. This module applies the core principles of data protection law, using the GDPR as the reference model and adding the main US developments, to the specific problems AI raises.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Apply transparency, lawful basis and purpose limitation to AI training and use, including the legitimate-interests test.
2. Explain data minimization and privacy by design for AI, and the privacy-enhancing techniques that support them.
3. Identify when a data protection impact assessment is needed and what it should cover for an AI system.
4. Describe how data-subject rights and the GDPR's rules on automated decisions apply to AI.
5. Recognize special-category data and biometric data in AI, including data inferred by a model.
6. Outline the US approach: state comprehensive privacy laws, California's rules on automated decision-making technology, and biometric privacy laws.

> **HOW TO USE THIS MODULE**
>
> Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. Why privacy law reaches AI

The GDPR applies to the processing of **personal data**: any information relating to an identified or identifiable person. It does not care whether a system is AI. But AI touches personal data in more places than most software:

Stage	Where personal data appears
Training	Datasets of customers, employees or people whose data was scraped from the web
The model	Models can memorize and later reveal training data, so the model itself may contain personal data
Inputs	Prompts, documents and records fed to the system in use
Outputs	Predictions, scores and generated text about people, including new inferences about them

Privacy law uses its own roles. A **controller** decides why and how personal data is processed; a **processor** processes it on the controller's behalf. These do not map neatly onto AI roles: a

deployer is usually the controller for its use; a vendor may be a processor for the deployer's data but a controller for its own model training.

> **EXAM TIP**
> Check which law's vocabulary a question uses. "Controller" and "processor" are privacy-law roles; "provider" and "deployer" are AI Act roles. One organization can hold one of each at the same time.

2. Lawful basis, transparency and purpose limitation

Every processing operation needs one of six **lawful bases** (GDPR Article 6). In AI the realistic candidates are:

Lawful basis	AI example	Watch out for
Consent	A user opts in to having their photos used to improve a feature	Must be free, specific, informed and withdrawable; hard to honor once data is in a trained model
Contract	Scoring a loan application the customer asked for	Covers what is necessary for the contract, not wider model training
Legal obligation	Fraud screening a law requires	The law must actually require the processing
Legitimate interests	Training a fraud model on transaction history	Requires the three-step test below; overridden where individuals' rights prevail

Public task and vital interests are the other two bases; they matter mainly for public bodies and emergencies.

The legitimate-interests test

Legitimate interests is the most common basis for AI development, and it requires three steps: (1) a **legitimate interest** that is lawful, clear and real; (2) **necessity**: the processing is needed for that interest and less intrusive means would not work; (3) a **balancing test**: the interest is not overridden by individuals' rights, considering their reasonable expectations, the data's nature and the safeguards in place. The European Data Protection Board's Opinion 28/2024 confirmed that this basis can be used for developing and deploying AI models, case by case, and that scraping public web data does not by itself meet people's reasonable expectations.

Transparency

People must be told what is processed, why, on what basis and for how long (Articles 12 to 14). Where solely automated decisions with significant effects are made, they must also receive **meaningful information about the logic involved** and the likely consequences. Where data was collected indirectly, such as by scraping, the duty to inform still applies, with limited exceptions.

Purpose limitation

Data collected for one purpose may not be reused for an incompatible one. Using customer-service transcripts to train a sales-targeting model is a new purpose that needs a **compatibility assessment** (Article 6(4)): the link between purposes, the context of collection, the nature of the data, the consequences for people and the safeguards. If it is incompatible, a fresh lawful basis is needed.

3. Data minimization and privacy by design

Data minimization requires personal data to be adequate, relevant and limited to what is necessary (Article 5(1)(c)). This sits in obvious tension with AI's appetite for data. The answer is not "no data" but justified data: define the purpose first, then show why each feature and each record is needed.

Data protection by design and by default (Article 25) requires the controller to build protections into systems from the start, and to process by default the minimum necessary. For AI, this means privacy decisions belong at the plan-and-design stage, not at launch.

Technique	What it does	Trade-off
Pseudonymization	Replaces identifiers with codes, keeping the key separate	Data stays personal data under the GDPR
Anonymization	Removes the ability to identify anyone, by any reasonably likely means	Hard to achieve in rich datasets; failure means the GDPR still applies
Synthetic data	Generates artificial records with similar statistical properties	May leak real records or miss real-world patterns
Differential privacy	Adds calibrated noise so no individual's presence can be inferred	Reduces accuracy, especially for small groups
Federated learning	Trains on data where it sits, sharing model updates rather than raw data	Updates can still leak information; complex to run

WATCH THE DISTINCTION

Pseudonymized data is still personal data. Anonymized data is not, but the bar is high, and a trained model can itself be personal data if information about individuals can be extracted from it. The EDPB treats a model's anonymity as a case-by-case question.

4. Controller obligations: assessments, rights and automated decisions

Data protection impact assessments

A **DPIA** is required where processing is likely to result in a high risk to people (Article 35), and specifically for systematic and extensive evaluation of people based on automated processing that produces legal or similarly significant effects, large-scale processing of special-category data, and systematic large-scale monitoring of public areas. Regulators' guidance treats new technologies, profiling, automated decisions and vulnerable data subjects as risk factors, so many AI uses will need one. A DPIA describes the processing and its purpose, assesses necessity and proportionality, assesses risks to people and sets out the measures to address them. Where high risk remains, the controller must consult the supervisory authority before processing.

Data-subject rights

Right	The AI difficulty
Access (Art. 15)	Locating a person's data across training sets and logs; explaining automated decisions
Rectification (Art. 16)	Correcting an inaccurate inference, or output a generative model keeps producing
Erasure (Art. 17)	Removing a person's influence from a trained model may mean retraining; "machine unlearning" is immature
Objection (Art. 21)	Stopping processing based on legitimate interests, including use in training, unless compelling grounds override

Automated decision-making (Article 22)

People have the right not to be subject to a decision **based solely on automated processing**, including profiling, that produces **legal or similarly significant effects** (for example refusing credit or a job). Such decisions are allowed only where necessary for a contract, authorized by law, or based on explicit consent, and then with safeguards: at least the right to obtain human intervention, to express a view and to contest the decision.

- **Human involvement must be meaningful.** A person who routinely rubber-stamps the system's output does not make the decision any less "solely automated".
- **Scores can be decisions.** In SCHUFA (C-634/21, 2023) the Court of Justice held that a credit score can itself be an automated decision when a lender relies on it heavily in deciding.
- **Explanations must be intelligible.** In Dun & Bradstreet Austria (C-203/22, 2025) the Court held that people are entitled to an explanation of the procedure and principles actually applied, not just a complex formula.

> **EXAM TIP**
> Article 22 applies to decisions that are solely automated and have legal or similarly significant effects. A recommendation that a person genuinely weighs is outside it, but the other GDPR duties still apply.

5. Transfers and breaches

AI services are often hosted outside the organization's country. Personal data leaving the European Economic Area needs a **transfer mechanism** (Chapter V): an adequacy decision (such as the EU–US Data Privacy Framework for certified US companies), standard contractual clauses or binding corporate rules, usually with a transfer impact assessment. Ask AI vendors where data is processed, including by sub-processors and model providers.

A **personal data breach** must be reported to the supervisory authority within **72 hours** of becoming aware of it, unless it is unlikely to result in a risk to people (Article 33), and to the people affected without undue delay where the risk is high (Article 34). AI adds new breach routes: a model revealing training data, a prompt injection that exfiltrates records, or staff pasting personal data into an unapproved tool.

6. Sensitive data and biometrics

Special categories of personal data (Article 9) are racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, genetic data, biometric data used to uniquely identify a person, health data, and data about sex life or sexual orientation. Processing them is prohibited unless an Article 6 lawful basis **and** an Article 9 condition apply, such as explicit consent, employment-law obligations or substantial public interest under law.

Inferred sensitive data

AI can infer sensitive traits from ordinary data: health from purchases, orientation from social connections, ethnicity from names and postcodes. The Court of Justice has held that data which reveals special-category information indirectly can itself be special-category data (OT, C-184/20, 2022). A model that infers sensitive traits brings Article 9 into play even if no sensitive data was collected.

Biometrics

Biometric data used for **verification** (are you who you claim to be?) and **identification** (who is this, among many?) carry different risks, with identification at scale the more intrusive. Biometric systems attract overlapping rules:

- the GDPR's special-category rules where biometric data is used to identify people;
- EU AI Act prohibitions, in force since 2 February 2025, including untargeted scraping of facial images to build recognition databases, emotion recognition in workplaces and schools (with medical and safety exceptions), and biometric categorization to infer sensitive traits;
- US state biometric laws, notably Illinois' Biometric Information Privacy Act (BIPA), which requires written notice and consent and gives individuals a private right of action, and similar laws in Texas and Washington.

7. The United States picture

The US has no comprehensive federal privacy law. Instead, roughly twenty states have passed comprehensive consumer privacy laws, and most share features that matter for AI:

- a right to **opt out of profiling** in furtherance of decisions that produce legal or similarly significant effects;
- **data protection assessments** for high-risk processing, including such profiling;
- **opt-in consent** for sensitive data in most states (California instead gives a right to limit its use).

California's ADMT rules

Regulations under the California Consumer Privacy Act add specific duties for businesses that use **automated decision-making technology** (ADMT) to make **significant decisions** about consumers: decisions that result in the provision or denial of financial or lending services, housing, education enrollment or opportunities, employment or independent-contracting opportunities or compensation, or healthcare services. They require:

- a **pre-use notice** explaining the purpose, how the ADMT works, the consumer's rights and the role of human involvement;
- a **right to opt out**, with exceptions: for example where the business offers an appeal to a human reviewer with authority to overturn the decision, or where ADMT is used solely for admission, hiring, work allocation or compensation decisions, works as intended for that purpose and does not discriminate;
- a **right to access** information about the business's use of ADMT for that consumer: its purpose, its logic, the outcome and how the output was used in the decision;
- **risk assessments** before processing that presents significant risk, including using ADMT for significant decisions and using personal information to train ADMT, with summary information submitted to the California Privacy Protection Agency.

CURRENCY NOTE (AS OF NOVEMBER 2026)

California: the regulations took effect on 1 January 2026. Businesses already using ADMT for significant decisions must comply by 1 January 2027. Risk assessments apply to new processing from 1 January 2026; those for processing already under way must be completed by 31 December 2027, and summary information for assessments in 2026 and 2027 is due to the agency by 1 April 2028. **EU:** the European Commission's November 2025 digital omnibus proposed GDPR amendments touching AI training, the definition of personal data and sensitive data. The Council agreed its negotiating position on 21 October 2026; the Parliament has not yet. The GDPR as described in this module is unchanged until amendments are adopted. Colorado's AI law and other AI-specific state laws are covered in Module 8.

Key terms

Term	Meaning
Personal data	Any information relating to an identified or identifiable living person.
Controller / processor	The party deciding why and how data is processed / the party processing on its behalf.
Lawful basis	One of six legal grounds for processing under GDPR Article 6.
Legitimate-interests test	Purpose, necessity and balancing: the three steps to rely on legitimate interests.
Purpose limitation	Data may not be reused for purposes incompatible with those it was collected for.
Data minimization	Personal data must be adequate, relevant and limited to what is necessary.
DPIA	Data protection impact assessment, required where processing is likely to be high-risk.
Article 22	GDPR right not to be subject to solely automated decisions with legal or similarly significant effects.
Special-category data	Sensitive data under GDPR Article 9, including health, biometric and ethnic-origin data.
ADMT	Automated decision-making technology, regulated in California under the CCPA.

Key takeaways

- Privacy law applies at every stage: training data, the model, inputs and outputs.
- Legitimate interests is the usual basis for AI development and needs purpose, necessity and balancing.
- Reusing data to train AI is often a new purpose: assess compatibility or find a new basis.
- Pseudonymized data is still personal; true anonymization is hard, and models can hold personal data.
- Article 22 covers solely automated, significant decisions; token human review does not take a decision out of it.
- Inferred sensitive traits are special-category data; biometrics face GDPR, AI Act and US state rules.
- In the US, state privacy laws give profiling opt-outs and assessments; California adds ADMT rules from 2027.

Knowledge check

8 questions. Choose the single best answer for each.

1. A bank wants to train a fraud model on its customers' transaction history and relies on legitimate interests. What must it show beyond having a legitimate interest?

- **A.** That customers gave explicit consent to training
- **B.** That it is necessary and not overridden by their rights
- **C.** That the model will be run outside the EU
- **D.** That the data is fully anonymized before any training starts

2. A telecom company wants to use customer-service call transcripts to train a model that targets upsell offers. What is the key privacy question?

- **A.** Whether reuse is compatible with the original purpose
- **B.** Whether the transcripts are stored in the cloud
- **C.** Whether the model is a large language model
- **D.** Whether the sales team has been trained to use the model

3. A team replaces customer names with random codes before training, keeping the key in a separate system. How does the GDPR treat the training data?

- **A.** As anonymous data that falls outside the GDPR
- **B.** As special-category data under Article 9
- **C.** As personal data, since it is pseudonymized
- **D.** As public data, since names were removed

4. Which approach best reflects data protection by design for a new AI system?

- **A.** Reviewing privacy once the model is ready for launch
- **B.** Collecting all available data in case it proves useful
- **C.** Letting the vendor decide which privacy measures apply
- **D.** Settling purpose and data needs at the design stage

5. A lender's model scores applicants, and loan officers approve nearly every score without looking further. Why may Article 22 still apply?

- **A.** Because credit data is special-category data
- **B.** Because the officers' review is not meaningful
- **C.** Because all lending models need explicit consent
- **D.** Because the model was trained outside the EU

6. An insurer plans an AI system that will evaluate applicants and set their premiums automatically. What does the GDPR require before processing begins?

- **A.** A data protection impact assessment
- **B.** Approval from the supervisory authority
- **C.** Registration in an EU AI database
- **D.** Explicit consent from every applicant

7. A retailer's model infers which shoppers are likely pregnant from their purchases. How should the inferences be treated?

- **A.** As ordinary personal data, since none was collected
- **B.** As anonymous data, since no names are used
- **C.** As outside the GDPR, since they are predictions
- **D.** As special-category data, since they reveal health

8. A company in Illinois wants to use employees' fingerprints for timekeeping. Under BIPA, what must it obtain first?

- **A.** A data protection impact assessment
- **B.** Approval from the Illinois attorney general's office
- **C.** Written notice and consent from each employee
- **D.** A certificate from an accredited auditor

Answers and explanations

#	Key	Why	Item ID · BoK
1	B	The legitimate-interests test has three steps: a legitimate interest, necessity, and a balancing test against individuals' rights and expectations.	FN-M6-01 · II.A.1
2	A	Purpose limitation: reuse for a new purpose needs a compatibility assessment, or else a fresh lawful basis.	FN-M6-02 · II.A.1
3	C	Pseudonymized data can be re-linked using the key, so it remains personal data. Anonymization requires that no one can reasonably be identified.	FN-M6-03 · II.A.2
4	D	Article 25 requires protections to be built in from the start, with the minimum necessary data processed by default.	FN-M6-04 · II.A.2
5	B	Token human review does not take a decision out of "solely automated". In SCHUFA, a heavily relied-on score was itself treated as an automated decision.	FN-M6-05 · II.A.3
6	A	Systematic, extensive automated evaluation with significant effects is a listed DPIA trigger. Prior consultation is needed only if high risk remains after mitigation.	FN-M6-06 · II.A.3
7	D	Data that reveals sensitive information indirectly, including inferences, can be special-category data, bringing Article 9 conditions into play.	FN-M6-07 · II.A.4
8	C	BIPA requires written notice and a written release before collecting biometric identifiers, and gives individuals a private right of action.	FN-M6-08 · II.A.4

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 7

Other Existing Law: IP, Discrimination, Consumer Protection and Liability

Domain II of the AIGP Body of Knowledge · II.B.1 to II.B.4 · About 45 minutes

Long before AI-specific laws arrived, AI systems were already subject to the laws that govern everything else an organization does. Copyright decides what can be used to train a model and who owns what it produces. Anti-discrimination law applies to a hiring algorithm exactly as it applies to a hiring manager. Consumer protection law punishes misleading claims about AI and unfair uses of it. Product liability decides who pays when an AI-enabled product causes harm. This module covers those four bodies of law as they apply to AI.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Explain the main intellectual-property questions AI raises: training on protected works, ownership of outputs, infringing outputs, trade secrets and likeness.
2. Apply disparate-treatment and disparate-impact concepts to AI in employment, credit, housing and insurance.
3. Describe how consumer protection law reaches deceptive AI claims and unfair AI practices, and the remedies regulators use.
4. Explain how product liability applies to AI, including the EU's revised Product Liability Directive.
5. Recognize that a deployer remains liable under these laws for a vendor's tool it chooses to use.

HOW TO USE THIS MODULE

Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. Existing law applies to AI

There is no AI exception to existing law. Regulators have said so plainly, and courts have treated AI systems as tools whose use is judged by the rules that already govern the activity. For governance, this has two consequences:

- **Legal review cannot wait for AI-specific law.** A use case with no AI-specific rule may still be unlawful under copyright, discrimination, consumer or liability law.
- **Using a vendor's tool does not move liability to the vendor.** An employer that screens applicants with a biased tool is answerable for the discrimination; a company whose chatbot misleads customers answers for what it said.

Body of law	The AI question it answers
Intellectual property	What may be used to train models? Who owns outputs? When does output infringe?
Non-discrimination	Is a system treating people differently, or producing unjustified disparities, in protected areas?
Consumer protection	Are claims about AI truthful? Is the use of AI unfair to consumers?
Product liability	Who pays when an AI-enabled product causes injury or loss?

2. Intellectual property

Training on protected works

Training a model involves copying works, so it engages copyright. Jurisdictions answer differently:

- **United States:** whether training is **fair use** is being decided case by case. Courts weigh the purpose of the use (including whether it is transformative), the nature of the work, how much is used, and the effect on the market for the original. Early decisions have gone both ways: copying used to build a directly competing product has been held not to be fair use (*Thomson Reuters v. Ross*, 2025), while other 2025 rulings treated training on lawfully acquired books as fair use but left the acquisition of pirated copies exposed.
- **European Union:** text-and-data-mining exceptions in the Copyright in the Digital Single Market Directive permit mining for research, and for other purposes **unless the rightsholder has opted out** in a machine-readable way. The EU AI Act requires general-purpose model providers to have a copyright policy that respects those opt-outs and to publish a summary of training content (Module 8).

Ownership of outputs

Copyright protects human authorship. The US Copyright Office and the courts (*Thaler v. Perlmutter*, 2025) have refused protection for material generated by AI without sufficient human creative control; prompts alone are not enough, but human selection, arrangement or modification of AI output can be protected. Patent offices and courts in the US and UK have likewise held that an inventor must be a natural person, while inventions made with AI assistance can still be patented by the human inventor.

Infringing outputs, trade secrets and likeness

Issue	What to watch
Infringing output	Generated content that reproduces or closely imitates a protected work, character or style can infringe. Vendor IP indemnities and output filters reduce, but do not remove, the risk.
Trade secrets	Confidential information entered into an external AI tool may lose trade-secret protection if it is no longer kept secret. Models and training data can themselves be trade secrets.
Trademarks	Generated content can misuse marks or imply false endorsement.
Likeness and voice	Right-of-publicity laws protect people's names, images and voices; some US states have added rules aimed at AI voice and likeness clones.

EXAM TIP

Separate the input question from the output question. "Can we train on this?" is about rights in the training data. "Can we protect or safely use what the model produces?" is about authorship and infringement. Options often answer the wrong one.

3. Non-discrimination law

Anti-discrimination law protects people from unfair treatment based on protected characteristics such as race, sex, age, disability and religion, in areas including employment, credit, housing and insurance. Two theories of liability matter for AI:

Theory	What it means	AI example
Disparate treatment	Treating people differently because of a protected characteristic, intentionally	A system set to auto-reject applicants above a certain age
Disparate impact	A neutral practice that disproportionately harms a protected group and is not justified by business necessity, or where a less discriminatory alternative exists	A screening model that rejects far more women than men because of features that act as proxies

Disparate impact is the bigger AI risk, because bias usually enters through data and proxies rather than intent. Removing the protected characteristic from the inputs does not remove the risk.

By sector (United States)

Area	Key law	AI-specific point
Employment	Title VII, the Age Discrimination in Employment Act, the Americans with Disabilities Act	Employers are responsible for tools they use; the four-fifths rule is a common screen for adverse impact; tools must not screen out people with disabilities who could do the job with reasonable accommodation
Credit	Equal Credit Opportunity Act (Regulation B), Fair Credit Reporting Act	Lenders must give applicants the specific principal reasons for adverse action, even when a complex model made the decision
Housing	Fair Housing Act	Tenant-screening and advertising-delivery algorithms have drawn enforcement
Insurance	State unfair-discrimination laws; insurance regulators' AI rules	Some states require insurers to test models and external data for unfair discrimination

Some laws target AI in hiring specifically, such as New York City's Local Law 144, which requires annual independent bias audits of automated employment decision tools and notice to

candidates. In the EU, equality directives, the GDPR and the AI Act's high-risk rules for employment and credit all apply.

> **CURRENCY NOTE (AS OF NOVEMBER 2026)**
> US federal enforcement priorities have shifted: a 2025 executive order directed federal agencies to deprioritize disparate-impact enforcement, and some agency AI guidance was withdrawn. The statutes themselves are unchanged, private lawsuits continue (for example *Mobley v. Workday*, in which a court allowed claims against an AI screening vendor to proceed), and state laws and regulators remain active. Teach the legal mechanisms, which the exam tests, rather than the current enforcement climate.

4. Consumer protection

In the US, Section 5 of the Federal Trade Commission Act prohibits **unfair or deceptive acts or practices**, and every state has a similar law enforced by its attorney general. In the EU, the Unfair Commercial Practices Directive plays a similar role. These laws reach AI in two ways.

Deceptive claims about AI

Claims about what an AI product can do must be truthful and substantiated. Overstating capability or accuracy, claiming a product uses AI when it does not ("AI washing"), or hiding material limitations can all be deceptive. The FTC's 2024 enforcement sweep against deceptive AI claims, and a Texas attorney-general settlement over a healthcare AI vendor's accuracy claims, are examples.

Unfair AI practices

A practice is unfair if it causes substantial injury that consumers cannot reasonably avoid and that is not outweighed by benefits. The FTC's 2023 order against Rite Aid, which used facial recognition that misidentified shoppers as suspected shoplifters, banned the company from using facial recognition for five years and required a governance program.

Remedies to know

- **Algorithmic disgorgement:** the FTC has required companies to delete models and algorithms built on improperly obtained data (*Everalbum*, 2021; *WW/Kurbo*, 2022).
- **Conduct bans** on using a technology for a period.
- **Mandated programs:** risk assessments, testing, monitoring and independent assessments imposed by order.

> **WATCH THE DISTINCTION**
> A business is responsible for what its AI tells customers. In *Moffatt v. Air Canada* (2024), a Canadian tribunal held the airline liable for its website chatbot's wrong statement about bereavement fares, rejecting the argument that the chatbot was a separate entity.

5. Product liability

Product liability makes manufacturers and sellers answer for injuries caused by defective products. The usual categories are **design defects**, **manufacturing defects** and **warning (or instruction) defects**. Claims may be brought in strict liability, where fault need not be proven, or in negligence, where the claimant must show a failure of reasonable care.

United States

A central question is whether software, including AI, is a "product" at all, or a service outside strict liability. Courts have split. In 2025 a federal court allowed product-liability claims against a companion-chatbot developer to proceed (*Garcia v. Character Technologies*). Negligence claims remain available either way.

European Union

The revised **Product Liability Directive** (Directive (EU) 2024/2853) modernizes strict liability for the digital age:

- **software, including AI systems, is expressly a product**;
- manufacturers can be liable for defects that arise after sale while they still control the product, including through updates or a failure to supply security updates;
- courts can order disclosure of evidence and **presume defectiveness or causation** in certain cases, easing the burden on claimants facing complex technology;
- recoverable damage includes psychological harm and the destruction or corruption of data.

> **CURRENCY NOTE (DECEMBER 2026)**
>
> The revised Directive applies to products placed on the market or put into service **after 9 December 2026**, the deadline for member states to transpose it; products placed on the market earlier remain under the 1985 Directive. Not every member state transposed on time, so check national law in the country concerned. The Commission withdrew its separate proposed AI Liability Directive in 2025, so fault-based AI claims remain a matter of national law.

Governance response

Liability law rewards the same controls as good governance: documented design choices and testing, clear instructions and warnings, monitoring after release, timely updates, and records that show reasonable care was taken.

Key terms

Term	Meaning
Fair use	US copyright defense weighing purpose, nature, amount and market effect of a use.
Text-and-data-mining exception	EU copyright exception permitting mining unless the rightsholder opts out (for non-research uses).
Human authorship	The requirement that copyright protects works with sufficient human creative contribution.
Disparate treatment	Intentional different treatment because of a protected characteristic.
Disparate impact	A neutral practice that disproportionately harms a protected group without adequate justification.
Four-fifths rule	Rule of thumb: a selection rate for a group below 80% of the highest group's rate suggests adverse impact.
Adverse action notice	Under US credit law, the statement of specific reasons for denying or worsening credit terms.
Unfair or deceptive practices	Conduct prohibited by FTC Act Section 5 and state equivalents.
Algorithmic disgorgement	A remedy requiring deletion of models built on improperly obtained data.
Strict liability	Liability for a defective product without proof of fault.

Key takeaways

- Existing law applies to AI now; there is no AI exception, and using a vendor's tool does not transfer liability.
- Training-data rights and output ownership are separate questions; copyright protects human authorship.
- Disparate impact is the main discrimination risk in AI, and removing protected attributes does not cure it.
- Credit decisions need specific reasons for adverse action, whatever the model's complexity.
- Consumer protection law reaches false AI claims and unfair AI uses; remedies include deleting models.
- In the EU, software and AI are now expressly products for strict liability, for products placed on the market after 9 December 2026.

Knowledge check

8 questions. Choose the single best answer for each.

1. A company wants to know whether it can claim copyright in marketing images an AI tool generated from short prompts with no further editing. What is the most likely answer under US law?

- **A.** Yes, because the company paid for the AI tool
- **B.** No, because there is no sufficient human authorship
- **C.** Yes, if the company registers the images promptly
- **D.** No, because AI images are in the public domain by statute

2. An AI developer wants to train a commercial model on web content in the EU. Under the text-and-data-mining exception, what must it respect?

- **A.** Rightsholders' machine-readable opt-outs
- **B.** A ban on mining any copyrighted work at all
- **C.** The need for a licence from every author
- **D.** A requirement to publish all training data

3. A lender's model denies an applicant credit. The lender says the model is too complex to give reasons. What does US credit law require?

- **A.** No reasons, since the decision was automated
- **B.** Disclosure of the model's full source code to the applicant
- **C.** The specific principal reasons for the denial
- **D.** A human re-decision before any notice is sent

4. A screening tool does not see applicants' sex, but it rejects women at a far higher rate because of proxy features. Which theory of liability fits best?

- **A.** Disparate treatment, since the tool intended it
- **B.** No liability, since sex was not an input
- **C.** Product liability for a design defect
- **D.** Disparate impact on a protected group

5. A start-up markets its tool as "AI-powered fraud detection with 99% accuracy", but it uses simple rules and has not been tested. Which law is most directly engaged?

- **A.** Consumer protection law on deceptive claims
- **B.** Copyright law on the tool's source code
- **C.** Anti-discrimination law in consumer credit decisions
- **D.** Product liability law in negligence

6. The FTC finds a company trained a model on photos collected without proper consent. Which remedy targets the model itself?

- **A.** A civil penalty paid to the US Treasury
- **B.** Algorithmic disgorgement of the model
- **C.** A public apology to affected users
- **D.** Registration in a federal AI database

7. Under the EU's revised Product Liability Directive, how is AI software treated?

- **A.** As a service outside product liability
- **B.** As a product if sold on physical media alone
- **C.** As a product subject to strict liability
- **D.** As outside the Directive if open source

8. A device maker stops supplying security updates for its AI-enabled home robot, and an exploit later causes injury. Why might the maker be liable under the revised EU Directive?

- **A.** Because any injury caused by a robot creates liability
- **B.** Because the user's own data was involved
- **C.** Because the robot was manufactured in the EU
- **D.** Because it failed to update a product it controlled

Answers and explanations

#	Key	Why	Item ID · BoK
1	B	US copyright requires human authorship; prompts alone are not enough. Human selection, arrangement or editing of output can be protected.	FN-M7-01 · II.B.1
2	A	For non-research mining, the EU exception applies unless the rightsholder has reserved its rights in a machine-readable way.	FN-M7-02 · II.B.1
3	C	ECOA and Regulation B require specific principal reasons for adverse action. Model complexity is not an excuse.	FN-M7-03 · II.B.2
4	D	A neutral practice with an unjustified disproportionate effect on a protected group is disparate impact. Removing the attribute does not prevent it.	FN-M7-04 · II.B.2
5	A	Unsubstantiated capability and accuracy claims, and claiming AI that is not there, are deceptive practices.	FN-M7-05 · II.B.3
6	B	Algorithmic disgorgement requires deletion of models and algorithms built on improperly obtained data.	FN-M7-06 · II.B.3
7	C	The revised Directive expressly treats software, including AI systems, as a product, so strict liability can apply.	FN-M7-07 · II.B.4
8	D	The revised Directive covers defects arising after sale while the manufacturer controls the product, including failure to supply security updates.	FN-M7-08 · II.B.4

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 8

The EU AI Act and Other AI-Specific Laws

Domain II of the AIGP Body of Knowledge · II.C.1 to II.C.6 · About 55 minutes

The EU AI Act is the world's first comprehensive AI law and the most heavily tested AI-specific law on the AIGP exam. This module explains its risk-based structure, what it requires of high-risk systems, how obligations divide by role, the separate regime for general-purpose AI models, and how it is enforced. It closes with the main AI-specific laws elsewhere, especially in US states. Much of this area is changing quickly, so every date carries an "as of" note.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Classify AI uses into the EU AI Act's risk tiers: prohibited, high-risk, transparency obligations and minimal risk.
2. List the requirements for high-risk AI systems, including risk management, data governance, documentation, logging, human oversight and quality management.
3. Distinguish the obligations of providers, deployers, importers and distributors.
4. Describe the obligations for general-purpose AI models, including those with systemic risk.
5. Explain enforcement and the penalty tiers.
6. Outline the main AI-specific laws in US states and other countries.

> **HOW TO USE THIS MODULE**
>
> Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. Shape and timeline of the EU AI Act

The AI Act (Regulation (EU) 2024/1689) entered into force on 1 August 2024. It is a product-safety-style regulation: obligations attach to AI systems and models according to their risk, and to the operators (Module 3) who place them on the market or use them. It applies to providers placing systems on the EU market wherever they are based, to deployers in the EU, and to providers and deployers outside the EU where the system's output is used in the EU.

Date	What applies
2 Feb 2025	Prohibited practices; AI literacy (Article 4)
2 Aug 2025	General-purpose AI model obligations; governance bodies; penalties
2 Aug 2026	General application, including transparency duties (Article 50) and Commission enforcement powers for GPAI
2 Dec 2026	Two new prohibited practices added by the Omnibus; end of the grace period for marking AI-generated content, for systems already on the market
2 Dec 2027	High-risk obligations for Annex III systems
2 Aug 2028	High-risk obligations for Annex I systems (AI in regulated products)

CURRENCY NOTE (AS OF NOVEMBER 2026)

The Digital Omnibus on AI (Regulation (EU) 2026/1744), in force since 27 July 2026, moved the high-risk dates from **2 August 2026** (Annex III) and **2 August 2027** (Annex I) to the dates above, softened the AI-literacy duty, added two prohibited practices and widened the use of sensitive data for bias detection. The exam blueprint predates these changes and may expect the original dates.

2. The risk tiers

Tier	What falls in it	Consequence
Prohibited (Article 5)	Practices with unacceptable risk	Banned
High-risk (Article 6)	Safety components of regulated products (Annex I); uses listed in Annex III	Full requirements, conformity assessment, registration
Transparency obligations (Article 50)	Systems that interact with people, generate content, recognize emotions or categorize biometrically, or produce deepfakes	Disclosure and marking duties
Minimal risk	Everything else, such as spam filters or game AI	No new obligations; voluntary codes encouraged

Prohibited practices

The main prohibitions cover AI that:

- uses subliminal, manipulative or deceptive techniques that materially distort behavior and cause significant harm;
- exploits vulnerabilities due to age, disability or social or economic situation;
- performs social scoring that leads to detrimental or disproportionate treatment;
- predicts a person's risk of committing a crime based solely on profiling or personality traits;
- builds facial-recognition databases by untargeted scraping of images;
- infers emotions in workplaces and education institutions, except for medical or safety reasons;
- categorizes people biometrically to infer sensitive traits such as race, political opinions or sexual orientation;
- performs real-time remote biometric identification in public spaces for law enforcement, subject to narrow exceptions.

The Digital Omnibus added two bans, applying from 2 December 2026: AI that generates or manipulates realistic intimate or sexually explicit images, video or audio of an identifiable person without their explicit consent, and AI that generates child sexual abuse material. For providers,

the ban covers systems where such output is the intended purpose or a reasonably foreseeable result they do not reasonably guard against.

High-risk systems

A system is high-risk if it is a safety component of, or is itself, a product covered by EU product legislation in **Annex I** that requires third-party conformity assessment (for example machinery, medical devices, toys), or if it falls in an **Annex III** area:

Annex III area	Examples
Biometrics	Remote biometric identification, biometric categorization, emotion recognition (where not prohibited)
Critical infrastructure	Safety components in managing traffic, water, gas, heating or electricity
Education	Admission decisions, grading, proctoring
Employment	Recruitment and selection, promotion and termination decisions, task allocation, performance monitoring
Essential services	Eligibility for public benefits, creditworthiness, life and health insurance pricing, emergency-call triage
Law enforcement	Evidence reliability, risk of victimization, profiling in investigations
Migration and borders	Risk assessment of travelers, examining asylum and visa applications
Justice and democracy	Assisting judicial decisions, influencing elections

> **EXAM TIP**
> Article 6(3) lets a provider conclude that an Annex III system is not high-risk if it poses no significant risk, for example because it performs a narrow procedural task or a preparatory task. That escape does not apply where the system profiles people. The provider must document its assessment.

3. Requirements for high-risk systems

High-risk systems must meet a set of requirements (Articles 8 to 15), which the provider must build in:

Requirement	What it means
Risk management system (Art. 9)	A continuous process across the life cycle to identify, evaluate and mitigate risks to health, safety and fundamental rights
Data and data governance (Art. 10)	Training, validation and testing data that is relevant, sufficiently representative and, as far as possible, free of errors; examination for bias

Requirement	What it means
Technical documentation (Art. 11)	Documentation showing compliance, drawn up before release and kept up to date (contents in Annex IV)
Record-keeping (Art. 12)	Automatic logging of events to allow traceability
Transparency to deployers (Art. 13)	Instructions for use covering purpose, accuracy, limitations, oversight measures and maintenance
Human oversight (Art. 14)	Design that lets people understand, monitor, interpret, override or stop the system, and guards against automation bias
Accuracy, robustness, cybersecurity (Art. 15)	Appropriate, declared levels of performance; resilience to errors and attacks

Providers must also run a **quality management system** (Art. 17), complete a **conformity assessment** before release (in most Annex III cases by internal control; for some biometric systems and Annex I products, with a notified body), sign an **EU declaration of conformity**, affix the **CE marking**, **register** the system in the EU database, run **post-market monitoring** and report **serious incidents**.

> **CURRENCY NOTE (AS OF NOVEMBER 2026)**
> Harmonized standards give a presumption of conformity. The first, EN 18286 on AI quality management systems, was cited in the Official Journal in October 2026; standards on risk management, logging, accuracy and datasets are still in preparation.

4. Obligations by role

Role	Main obligations for high-risk systems
Provider	Meet the requirements; quality management system; documentation; conformity assessment, declaration and CE marking; registration; post-market monitoring; serious-incident reporting; corrective action
Deployer (Art. 26)	Use in line with the instructions for use; assign human oversight to competent, trained people; ensure input data under its control is relevant and representative; monitor operation and inform the provider of risks; keep logs for at least six months; inform workers' representatives before using it at work; inform people that a high-risk system is used in decisions about them
Importer (Art. 23)	Before placing a non-EU provider's system on the market, verify conformity assessment, documentation, CE marking and an authorized representative; add its own name and contact details
Distributor (Art. 24)	Verify CE marking, the declaration of conformity and instructions; do not supply a system it believes is non-compliant

Two further deployer duties are frequently tested. A **fundamental rights impact assessment** (Art. 27) is required before first use from public bodies and private entities providing public services, and from deployers using high-risk systems for creditworthiness or life and health insurance pricing. And people affected by a decision based on a high-risk system's output have a **right to an explanation** of the system's role and the main elements of the decision (Art. 86).

Transparency obligations (Article 50)

Who	Duty
Providers of systems that interact with people	Design them so people are told they are dealing with AI, unless obvious
Providers of generative systems	Mark synthetic audio, image, video and text in a machine-readable, detectable way
Deployers of emotion recognition or biometric categorization	Inform the people exposed to it
Deployers of deepfakes	Disclose that the content is artificially generated or manipulated
Deployers publishing AI-generated text on matters of public interest	Disclose it, unless it has undergone human editorial review and someone holds editorial responsibility

5. General-purpose AI models

Providers of **general-purpose AI (GPAI) models** have their own obligations, separate from the risk tiers for systems (Articles 53 to 55):

- technical documentation for the AI Office and national authorities;
- information and documentation for downstream providers who build systems on the model;
- a policy to comply with EU copyright law, including respecting text-and-data-mining opt-outs;
- a sufficiently detailed public summary of the content used for training, using the Commission's template.

Open-source models released under a free licence are exempt from the first two duties, unless they pose systemic risk.

Systemic risk

A GPAI model has **systemic risk** if it has high-impact capabilities, presumed when training compute exceeds **10^25 (ten to the twenty-fifth power) floating-point operations**, or if the Commission designates it. Its provider must also evaluate the model, including adversarial testing; assess and mitigate systemic risks; track and report serious incidents; and ensure adequate cybersecurity.

The **GPAI Code of Practice**, published in July 2025, is a voluntary way for providers to show compliance. Models placed on the market before 2 August 2025 have until 2 August 2027 to comply.

> **WATCH THE DISTINCTION**
> Obligations for GPAI models fall on the model's provider. A company that builds a high-risk system on top of a GPAI model is the provider of that system and carries the high-risk obligations for it.

6. Enforcement and penalties

Each member state designates **market surveillance authorities** to enforce the Act for AI systems. The Commission's **AI Office** supervises GPAI models and, since the Omnibus, also has exclusive competence, with some exceptions, over AI systems built on a GPAI model by the same provider or undertaking, and over AI systems integrated into very large online platforms and search engines. The **European AI Board** coordinates national authorities.

Infringement	Maximum fine
Prohibited practices	EUR 35 million or 7% of worldwide annual turnover, whichever is higher
Most other obligations, including high-risk and transparency duties	EUR 15 million or 3%
Supplying incorrect or misleading information to authorities	EUR 7.5 million or 1%
GPAI model providers (imposed by the Commission)	EUR 15 million or 3%

For SMEs, and since the Omnibus small mid-caps, the cap is whichever of the two amounts is **lower**.

7. AI-specific laws beyond the EU

United States

There is no comprehensive federal AI law. Federal regulators apply existing law (Module 7), and executive-branch policy has shifted repeatedly, including attempts to limit state AI laws. AI-specific law is mostly made by states and cities:

Law	What it does
Colorado	The 2024 Colorado AI Act (SB 24-205) imposed a duty of reasonable care on developers and deployers of high-risk systems to avoid algorithmic discrimination, with impact assessments and notices. It was repealed before taking effect and replaced in 2026 by SB 26-189, a narrower transparency law for automated decision-making technology, effective 1 January 2027.
New York City Local Law 144	Requires bias audits of automated employment decision tools and notice to candidates.
California	CCPA rules on automated decision-making technology (Module 6), and a transparency law for developers of frontier models (SB 53).
Other states	Laws on AI disclosure (Utah), government and prohibited uses (Texas), AI in employment decisions (Illinois), deepfakes in elections, and insurers' use of AI.

Other jurisdictions

- **China** regulates specific technologies: algorithmic recommendation, deep synthesis, generative AI services and, since 2025, labeling of AI-generated content.
- **South Korea's** AI Basic Act took effect on 22 January 2026, with obligations for high-impact and generative AI.
- **Canada's** proposed federal AI law (AIDA) lapsed in 2025; **the UK** relies on existing regulators applying cross-sector principles.

> **CURRENCY NOTE (AS OF NOVEMBER 2026)**
>
> SB 26-189 takes effect on 1 January 2027. It gives consumers rights including meaningful human review of consequential decisions where commercially reasonable, or else an explanation of the factors relied on. Colorado's Department of Law adopted implementing rules on 19 November 2026; they take effect 20 days after publication. The exam may still expect SB 24-205's duty-of-care model, so know both shapes and why the law changed. State AI laws are being introduced and amended every session; check before teaching any specific requirement.

Key terms

Term	Meaning
Prohibited practice	An AI use banned outright under Article 5 of the AI Act.
High-risk AI system	A system in an Annex I product or an Annex III area, subject to the Act's full requirements.
Conformity assessment	The process showing a high-risk system meets the requirements before release.
CE marking	The mark showing a product conforms with EU requirements.
Instructions for use	The provider's information that lets deployers use a high-risk system properly.
FRIA	Fundamental rights impact assessment required of certain deployers before first use.
GPAI model	A general-purpose AI model capable of a wide range of tasks.
Systemic risk	Risk from high-impact GPAI models, presumed above 10^25 FLOPs of training compute.
AI Office	The Commission body supervising GPAI models and coordinating enforcement.
Digital Omnibus on AI	Regulation (EU) 2026/1744, which amended the AI Act, including its timeline.

Key takeaways

- The AI Act is risk-based: prohibited, high-risk, transparency obligations, minimal risk.
- High-risk means an Annex I product or an Annex III use; the Article 6(3) escape does not cover profiling.
- Providers build compliance in; deployers use systems properly, oversee them, keep logs and inform people.
- GPAI providers owe documentation, downstream information, a copyright policy and a training summary; systemic-risk models owe more.
- Fines reach EUR 35 million or 7% of turnover for prohibited practices.
- Outside the EU, AI-specific law is fragmented and fast-moving: know the shapes, check the dates.

Knowledge check

8 questions. Choose the single best answer for each.

1. A bank uses an AI system to assess the creditworthiness of consumers applying for loans. How does the EU AI Act classify it?

- **A.** As a prohibited social-scoring practice
- **B.** As a high-risk system listed in Annex III
- **C.** As minimal risk, since banks are regulated
- **D.** As a general-purpose AI model

2. A company plans to use AI to infer employees' emotions from webcam footage during meetings. What is the AI Act status of this use?

- **A.** Prohibited emotion recognition at work
- **B.** High-risk, subject to a conformity assessment
- **C.** Allowed with a transparency notice
- **D.** Allowed if employees give consent

3. Which requirement obliges a high-risk system's provider to ensure training data is relevant, sufficiently representative and examined for bias?

- **A.** Record-keeping under Article 12
- **B.** Human oversight under Article 14
- **C.** Data governance under Article 10
- **D.** Transparency obligations under Article 50

4. A provider designs its high-risk system so operators can monitor it, interpret its outputs and stop it. Which requirement is this?

- **A.** Accuracy and robustness
- **B.** A quality management system
- **C.** Post-market monitoring
- **D.** Human oversight by design

5. Which duty applies to every provider of a general-purpose AI model, including those without systemic risk?

- **A.** Adversarial testing of the model
- **B.** Reporting serious incidents to the AI Office
- **C.** A summary of the content used for training
- **D.** A fundamental rights impact assessment

6. What is the maximum fine for engaging in a prohibited AI practice under the EU AI Act?

- **A.** EUR 35 million or 7% of worldwide turnover
- **B.** EUR 15 million or 3% of worldwide turnover
- **C.** EUR 7.5 million or 1% of worldwide turnover
- **D.** EUR 20 million or 4% of worldwide turnover

7. A hospital in the EU uses a high-risk triage system bought from a vendor. Which obligation falls on the hospital as deployer?

- **A.** Carrying out the conformity assessment
- **B.** Assigning trained staff to oversee it
- **C.** Affixing the CE marking to the system
- **D.** Registering as the system's provider

8. A distributor learns that a high-risk system it is about to supply lacks CE marking. What should it do?

- **A.** Supply it and inform the deployer
- **B.** Add its own CE marking to the system
- **C.** Carry out the conformity assessment itself
- **D.** Withhold the system until it conforms

Answers and explanations

#	Key	Why	Item ID · BoK
1	B	Creditworthiness assessment of natural persons is an Annex III essential-services use, so the system is high-risk.	FN-M8-01 · II.C.1
2	A	Emotion recognition in the workplace is prohibited, except for medical or safety reasons. Consent does not lift the ban.	FN-M8-02 · II.C.1
3	C	Article 10 sets data and data-governance requirements for training, validation and testing data, including examination for possible biases.	FN-M8-03 · II.C.2
4	D	Article 14 requires high-risk systems to be designed for effective human oversight, including the ability to override or stop them.	FN-M8-04 · II.C.3
5	C	All GPAI providers must publish a training-content summary. Adversarial testing and incident reporting are extra duties for systemic-risk models; FRIAs are for certain deployers.	FN-M8-05 · II.C.4
6	A	Prohibited practices carry the top tier: EUR 35 million or 7%, whichever is higher (lower for SMEs). EUR 20 million or 4% is the GDPR's top tier.	FN-M8-06 · II.C.5
7	B	Deployers must assign human oversight to competent, trained people and use the system per its instructions. Conformity assessment and CE marking are provider duties.	FN-M8-07 · II.C.6
8	D	A distributor must verify CE marking and documentation and must not make available a system it considers non-compliant.	FN-M8-08 · II.C.6

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 9

Standards and Frameworks: OECD, NIST and ISO/IEC

Domain II of the AIGP Body of Knowledge · II.D.1 to II.D.3 · About 40 minutes

Laws say what must be achieved; standards and frameworks show how. Most organizations build their AI governance on one or more of three families: the OECD's principles and classification framework, the US NIST AI Risk Management Framework, and the ISO/IEC standards on AI concepts, management systems and impact assessment. This module explains what each is for, how each is structured, and how to choose and combine them.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Describe the OECD AI Principles, the OECD definition of an AI system and the OECD Framework for the Classification of AI Systems.
2. Explain the structure of the NIST AI Risk Management Framework: its trustworthiness characteristics, its four core functions, the Playbook and profiles.
3. Distinguish ISO/IEC 22989, 42001, 42005 and 23894, and explain what certification to ISO/IEC 42001 involves.
4. Choose and combine frameworks to fit an organization's needs, and explain how standards relate to legal compliance.

> **HOW TO USE THIS MODULE**
> Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. Why frameworks matter

Frameworks and standards turn principles and legal duties into repeatable practice. They give an organization a common vocabulary, a structure for its program, a way to show regulators, customers and auditors what it does, and, in some cases, a certificate.

Source	Type	Binding?	Main use
OECD AI Principles	Intergovernmental policy standard	No (political commitment by governments)	Shared values and definitions; policy alignment
NIST AI RMF	Voluntary risk framework	No	Flexible risk-management structure and vocabulary
ISO/IEC 42001	International management-system standard	No, but certifiable	Auditable program with third-party certification
EU harmonized standards	European standards cited under the AI Act	No, but compliance gives a presumption of conformity	Showing high-risk systems meet the AI Act

WATCH THE DISTINCTION

Following a voluntary framework is not the same as complying with the law. Certification to ISO/IEC 42001 does not by itself show compliance with the EU AI Act; only harmonized standards cited in the Official Journal give a legal presumption of conformity (Module 8).

2. OECD

The AI Principles

The OECD AI Principles, adopted in 2019 and updated in 2024, were the first intergovernmental AI standard and underpin the G20 AI principles. They have two parts.

Values-based principles	Recommendations to governments
Inclusive growth, sustainable development and well-being	Invest in AI research and development
Respect for the rule of law, human rights and democratic values, including fairness and privacy	Foster an inclusive AI-enabling ecosystem
Transparency and explainability	Enable an interoperable governance and policy environment
Robustness, security and safety	Build human capacity and prepare for labor-market transition
Accountability	Pursue international cooperation for trustworthy AI

The OECD's definition of an AI system, revised in 2023, is the basis of the EU AI Act's definition (Module 1).

The Classification Framework

The OECD Framework for the Classification of AI Systems (2022) helps policymakers and organizations characterize a system's risks by looking at five dimensions:

Dimension	Questions it raises
People and planet	Who is affected, and how? Which rights, users and stakeholders are involved?
Economic context	In which sector and business function is the system used? How critical is it?
Data and input	Where does the data come from, how is it collected, and what is its quality?
AI model	What type of model is it, and how was it built and trained?

Dimension	Questions it raises
Task and output	What does the system do, how autonomous is it, and what are its outputs used for?

The OECD also runs the AI Incidents and Hazards Monitor, which tracks reported AI incidents worldwide.

3. The NIST AI Risk Management Framework

NIST published the AI RMF 1.0 in January 2023. It is **voluntary**, **sector-agnostic** and **use-case agnostic**, intended for any organization that designs, develops, deploys or uses AI. Part 1 frames AI risk and defines trustworthy AI (Module 2's seven characteristics). Part 2 is the **Core**.

Function	Purpose	Examples of outcomes
Govern	Cultivates a culture of risk management; applies across all the other functions	Policies, accountability, roles, workforce diversity, third-party risk processes
Map	Establishes context and identifies risks	Intended purpose, context of use, affected people, system categorization
Measure	Analyzes, assesses and tracks identified risks	Metrics, testing for trustworthiness characteristics, tracking over time
Manage	Prioritizes and acts on risks	Risk responses, resources, incident response, decommissioning

Each function breaks down into **categories** and **subcategories**. The companion **Playbook** gives suggested actions, references and documentation for each subcategory; it is guidance, not a checklist.

Profiles

A **profile** tailors the framework to a setting. Use-case profiles apply it to a sector or application; **current and target profiles** compare where an organization is with where it wants to be; cross-sectoral profiles address a technology, such as NIST's **Generative AI Profile** (NIST AI 600-1, 2024), which lists risks specific to generative AI and suggested actions.

> **EXAM TIP**
> Govern is cross-cutting: it informs and runs through Map, Measure and Manage rather than being a step at the start. The functions are not strictly sequential.

CURRENCY NOTE (AS OF NOVEMBER 2026)

Version 1.0 is current. NIST has said a revision is under way, with no published date.

4. The ISO/IEC standards

ISO and IEC develop AI standards through their joint committee, ISO/IEC JTC 1/SC 42. Four are central:

Standard	What it covers	Certifiable?
ISO/IEC 22989 (2022)	AI concepts and terminology: definitions, life cycle stages, stakeholder roles such as AI provider, producer, customer, partner and subject	No
ISO/IEC 23894 (2023)	Guidance on AI risk management, adapting the ISO 31000 risk-management process	No
ISO/IEC 42001 (2023)	Requirements for an AI management system (AIMS)	Yes
ISO/IEC 42005 (2025)	Guidance on AI system impact assessment: assessing effects on individuals, groups and society	No

ISO/IEC 42001 in more detail

ISO/IEC 42001 follows the same harmonized structure as ISO/IEC 27001 and other management-system standards, so organizations can integrate it with existing programs. Its clauses follow the Plan-Do-Check-Act cycle:

- **Context (4):** scope, interested parties, the organization's role with respect to AI.
- **Leadership (5):** commitment, AI policy, roles and responsibilities.
- **Planning (6):** AI risk assessment and treatment, AI system impact assessment, objectives.
- **Support (7):** resources, competence, awareness, communication, documented information.
- **Operation (8):** carrying out the planned processes, assessments and treatments.
- **Performance evaluation (9):** monitoring, internal audit, management review.
- **Improvement (10):** nonconformity, corrective action, continual improvement.

Annex A lists reference controls (38 controls in nine areas, A.2 to A.10) covering areas such as AI policies, internal organization, resources, impact assessment, the AI system life cycle, data, information for interested parties, use of AI systems and third-party relationships. The organization selects controls and records its choices and justifications in a **Statement of Applicability**. Certification is by an accredited third-party auditor; ISO/IEC 42006 sets requirements for those certification bodies.

5. Choosing and combining

Need	Best starting point
Shared language and principles across an organization or with governments	OECD Principles and definition
A flexible, risk-based structure without certification	NIST AI RMF and Playbook
An auditable program and a certificate to show customers	ISO/IEC 42001
A structured impact assessment method	ISO/IEC 42005 (and the NIST Map function)
Legal presumption of conformity for EU high-risk systems	Harmonized standards cited under the AI Act

The frameworks are designed to fit together. NIST publishes crosswalks to ISO/IEC 42001 and other standards, and many organizations use NIST's vocabulary for risk work inside a 42001 management system. The governance question is not which single framework is right, but which combination meets the organization's legal, customer and internal needs without duplicate effort.

> **EXAM TIP**
> Match the framework to the need in the scenario. "Certification" or "third-party audit" points to ISO/IEC 42001. "Voluntary, flexible, Govern-Map-Measure-Manage" points to NIST. "Intergovernmental principles" points to the OECD. "Terminology" points to ISO/IEC 22989.

Key terms

Term	Meaning
OECD AI Principles	Intergovernmental principles for trustworthy AI, adopted 2019 and updated 2024.
OECD Classification Framework	Five dimensions for characterizing AI systems: people and planet, economic context, data and input, AI model, task and output.
NIST AI RMF	Voluntary US framework for managing AI risk, version 1.0 (2023).
Govern, Map, Measure, Manage	The four functions of the NIST AI RMF Core.
Playbook	NIST's companion guidance with suggested actions for each subcategory.
Profile	A tailored application of the NIST AI RMF to a sector, use case or technology.
ISO/IEC 22989	International standard on AI concepts and terminology.
ISO/IEC 42001	Certifiable requirements for an AI management system.
Statement of Applicability	The 42001 record of which Annex A controls apply and why.
ISO/IEC 42005	Guidance on AI system impact assessment.

Key takeaways

- Frameworks show how to govern; following one is not the same as complying with law.
- The OECD gives principles, the definition behind the EU AI Act, and a five-dimension classification framework.
- NIST's Core is Govern, Map, Measure, Manage, with Govern cutting across the rest; profiles tailor it.
- ISO/IEC 42001 is the certifiable AI management system; 22989 is terminology, 23894 risk guidance, 42005 impact assessment.
- Combine frameworks to fit legal, customer and internal needs.

Knowledge check

8 questions. Choose the single best answer for each.

1. Which of the following is one of the OECD's values-based AI principles?

- **A.** Accountability
- **B.** Mandatory certification
- **C.** Open source
- **D.** Licensing

2. A policy team wants to characterize an AI system by its affected people, sector, data, model and task. Which tool fits best?

- **A.** The NIST Generative AI Profile
- **B.** The OECD Classification Framework
- **C.** The ISO/IEC 42001 Annex A controls
- **D.** The EU AI Act conformity checklist

3. In the NIST AI RMF, which function is described as cross-cutting, informing the other three?

- **A.** Map
- **B.** Measure
- **C.** Manage
- **D.** Govern

4. A team is establishing the intended purpose, context of use and affected people for a new AI system under the NIST AI RMF. Which function is it performing?

- **A.** Govern
- **B.** Map
- **C.** Measure
- **D.** Manage

5. What is the NIST AI RMF Playbook?

- **A.** A mandatory checklist for federal contractors
- **B.** A certification scheme run by NIST
- **C.** Suggested actions for each subcategory
- **D.** A list of prohibited AI practices

6. A software company wants an independent certificate showing customers it runs a structured AI governance program. Which standard should it adopt?

- **A.** ISO/IEC 42001, AI management systems
- **B.** ISO/IEC 22989, AI concepts and terminology
- **C.** ISO/IEC 23894, AI risk management guidance
- **D.** ISO/IEC 42005, impact assessment

7. Under ISO/IEC 42001, where does an organization record which Annex A controls it has selected and why?

- **A.** In its AI system impact assessment report
- **B.** In the management review minutes
- **C.** In its Statement of Applicability
- **D.** In the certification body's audit report

8. Which standard provides guidance specifically on assessing an AI system's effects on individuals, groups and society?

- **A.** ISO/IEC 22989 on AI concepts and terms
- **B.** ISO/IEC 27001 on information security
- **C.** ISO/IEC 23894 on risk management
- **D.** ISO/IEC 42005 on impact assessment

Answers and explanations

#	Key	Why	Item ID · BoK
1	A	The five values-based principles are inclusive growth; rule of law and human rights; transparency and explainability; robustness, security and safety; and accountability.	FN-M9-01 · II.D.1
2	B	The OECD Framework for the Classification of AI Systems uses exactly these five dimensions.	FN-M9-02 · II.D.1
3	D	Govern cultivates the risk-management culture and runs through Map, Measure and Manage.	FN-M9-03 · II.D.2
4	B	Map establishes context and identifies risks, including purpose, context of use and affected people.	FN-M9-04 · II.D.2
5	C	The Playbook is voluntary companion guidance with suggested actions, references and documentation for each subcategory.	FN-M9-05 · II.D.2
6	A	ISO/IEC 42001 sets requirements for an AI management system and is the one that can be certified by an accredited third party.	FN-M9-06 · II.D.3
7	C	The Statement of Applicability records the controls chosen and the justification for including or excluding them.	FN-M9-07 · II.D.3
8	D	ISO/IEC 42005 (2025) gives guidance on AI system impact assessment. 23894 covers organizational AI risk management.	FN-M9-08 · II.D.3

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 10

Governing AI Design, Data and Testing

Domain III of the AIGP Body of Knowledge · III.A.1 to III.B.5 · About 55 minutes

Domain III of the AIGP Body of Knowledge asks how an organization governs AI it develops. This module covers the first two of its three competencies: designing and building a system, and governing the data, training and testing behind it. These are the stages where governance is cheapest and most effective, because a problem caught in design costs a meeting, while the same problem caught after release can cost a recall, a lawsuit or a regulator's order.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Define the business context and use case for an AI system, including whether AI is the right tool.
2. Plan and conduct an AI impact assessment, and coordinate it with legally required assessments.
3. Turn policies and ethical commitments into design requirements, and manage design and build risks.
4. Describe what to document during design and build, and why.
5. Set data governance requirements, and record data lineage and provenance.
6. Plan, perform and document training and testing, including fairness, robustness, security and explainability testing.

> **HOW TO USE THIS MODULE**
>
> Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. Business context and use case (III.A.1)

Governance starts before any data is touched, with a precise statement of what the system is for. A good use-case definition answers:

- **Problem and purpose:** what business problem is being solved, and what the system will do (predict, classify, recommend, generate, act).
- **Decision and stakes:** which decision the output informs or makes, how consequential it is for the people affected, and whether a person stays in the loop.
- **People:** who will use the system, and who will be affected by it without using it.
- **Success measures:** how the organization will know the system works, in business terms and in performance terms.
- **Constraints:** legal, contractual, ethical, budget and time limits.
- **Alternatives:** whether a simpler approach, such as rules, a checklist or better data, would solve the problem with less risk.

> **EXAM TIP**
>
> "Is AI the right tool?" is a legitimate governance answer. When a scenario describes a problem a fixed rule would solve, or a high-stakes decision with little data, the best response may be not to build an AI system at all.

The use case also fixes the system's **intended purpose**, which drives its risk tier (Module 4), its legal classification (Module 8) and the limits written into its documentation. Uses outside the intended purpose need their own review.

2. Impact assessment (III.A.2)

An **AI impact assessment** examines how the system could affect individuals, groups and society, and the organization itself. It is started at design, revisited at each gate and updated whenever the system or its context changes.

Step	What happens
Scope	Describe the system, its purpose, users, affected people and setting
Identify impacts	Consider effects on rights (privacy, non-discrimination, due process), safety, well-being, access to services, the environment and the organization
Assess	Rate each impact by severity, likelihood, scale and reversibility, giving weight to vulnerable groups
Consult	Involve affected stakeholders or their representatives, domain experts and internal functions
Mitigate	Decide design changes, safeguards, oversight and monitoring that reduce each impact
Decide and record	Record residual impacts, who accepted them and the conditions attached

One assessment, several purposes

Many assessments can apply to one system: a privacy DPIA (Module 6), an EU AI Act fundamental rights impact assessment for certain deployers, a data protection assessment under US state law, and the organization's own AI impact assessment. Good practice is to run one coordinated process whose output satisfies each requirement, rather than several disconnected exercises that reach different conclusions. ISO/IEC 42005 gives guidance on structuring the assessment (Module 9).

> **WATCH THE DISTINCTION**
>
> A **risk assessment** asks what could go wrong for the organization and how likely it is. An **impact assessment** asks what the system could do to people and society. A complete review needs both; exam options sometimes swap them.

3. Applying policies and managing design risks (III.A.3, III.A.4)

Design is where policy statements become engineering requirements. Each principle needs a concrete, testable translation:

Commitment	Design requirement it might become
Fairness	Named protected groups to test; prohibited proxy features; maximum acceptable disparity in error rates
Transparency and explainability	Users told AI is involved; reason codes for each decision; documentation for reviewers
Human oversight	Defined points where a person reviews or approves; ability to override and stop the system
Privacy	Minimum data; pseudonymization; retention limits; no sensitive inferences
Safety and reliability	Minimum performance thresholds; fallback when confidence is low; behavior limits
Security	Threat model; access controls on data and models; input and output filtering

Design and build risks

Common risks at this stage include an unclear or shifting purpose, unsuitable or unlawful data, choosing a model too opaque for the explanation the use requires, proxies for protected traits, over-reliance on a third-party component, and security weaknesses in the pipeline. Each is logged in a risk register with an owner, a rating and a response: **avoid** (change the design), **mitigate** (add a control), **transfer** (for example by contract or insurance) or **accept** (with sign-off at the right level).

Choosing a model

Model choice is a governance decision as well as a technical one. A simpler, interpretable model may be preferable where decisions must be explained, even at some cost in accuracy; a complex model may be justified where accuracy saves lives and explanations can be supplied by other means. Record the trade-off and who decided it (Module 2).

4. Documenting design and build (III.A.5)

Documentation is created as the work happens, not reconstructed afterwards. A design-and-build record typically includes:

- the use-case definition and intended purpose, including uses that are out of scope;

- requirements derived from policies, law and the impact assessment;
- architecture, model selection and the alternatives considered;
- a decision log of trade-offs, with who decided and why;
- the risk register and how each risk was treated;
- dependencies, including third-party models, data and tools.

This record serves several audiences at once: approvers at each gate, auditors, regulators (the EU AI Act requires technical documentation for high-risk systems, Module 8), engineers who maintain the system later, and the organization's lawyers if a decision is ever challenged.

> **EXAM TIP**
> If a question asks how an organization can show it took reasonable care, the answer usually involves contemporaneous documentation: records made at the time, showing what was considered and decided.

5. Data governance, lineage and provenance (III.B.1, III.B.2)

Data governance requirements

Requirement	What it asks
Lawfulness and rights	Is there a legal basis and the right (licence, consent, contract) to use the data for training?
Fitness for purpose	Is the data relevant to the task and the population the system will serve?
Quality	Is it accurate, complete, consistent and current enough? Are labels reliable?
Representativeness	Are all relevant groups present in sufficient numbers? Are there gaps that will produce weaker performance for some people?
Bias examination	Does the data reflect historical discrimination, measurement problems or proxies?
Minimization and protection	Is only necessary data used, with appropriate security, access control and retention?

Data used for **training**, **validation** (tuning the model) and **testing** (final evaluation) should be kept separate, so the test result is an honest estimate of real-world performance.

Lineage and provenance

Provenance records where data came from: its source, collection method, date, licence and any consent terms. **Lineage** records what happened to it afterwards: cleaning, labeling, filtering,

joining, augmentation and the version that was used to train each model. Together they let an organization answer, for any model version: what data trained it, were we entitled to use it, and what changed since the last version? They are also what makes it possible to honor a deletion request or remove a tainted dataset.

Datasheets for datasets (or data cards) are a common standard format for recording a dataset's motivation, composition, collection process, preprocessing, intended uses and limitations.

6. Planning and performing training and testing (III.B.3, III.B.4)

Plan before you test

A test plan sets out, before testing starts: what will be tested, the metrics, the **acceptance criteria** (thresholds the system must meet), the test data and how it was obtained, who will test (with independence proportionate to the risk), and what happens if criteria are missed. Setting criteria in advance prevents results being rationalized after the fact.

Test type	What it checks	Typical methods
Performance	Accuracy, precision and recall, error rates, calibration	Held-out test sets, cross-validation, results broken down by subgroup
Fairness	Disparities in outcomes or errors between groups	Selection-rate comparisons, equal-error-rate checks, calibration by group
Robustness	Behavior under noisy, unusual or shifted inputs	Stress tests, edge cases, out-of-distribution data
Security	Resistance to attack	Adversarial examples, prompt-injection tests, data-extraction attempts
Explainability	Whether outputs can be explained to the required level	Feature-importance methods, counterfactual explanations, reviewer comprehension tests
Human factors	Whether operators understand and appropriately challenge outputs	Usability testing, oversight simulations
Generative output quality	Factual accuracy, harmful content, consistency	Evaluation sets, human rating, red-team prompts

Fairness metrics conflict

Common fairness measures include **demographic parity** (equal selection rates), **equal opportunity** (equal true-positive rates), **equalized odds** (equal true- and false-positive rates) and **calibration** (scores mean the same thing for each group). Where base rates differ between groups, these generally cannot all be satisfied at once. Choosing the measure is a policy

decision that depends on the context and the law, and should be made and recorded by the right people, not left to whoever writes the test.

Managing issues in training and testing

Issues found at this stage include overfitting, data leakage, label errors, unrepresentative test data, performance gaps between groups, and test results that depend on how the test was set up. Each is logged and resolved before release, or formally accepted with conditions. Retesting after a fix is part of the process.

7. Documenting training and testing (III.B.5)

The training and testing record should let a reviewer reproduce and judge the result:

- the model version, code version and training configuration;
- the exact datasets and versions used for training, validation and testing (linking to lineage records);
- the test plan, the acceptance criteria and the results, including subgroup results;
- issues found, how each was resolved or accepted, and by whom;
- known limitations and conditions on use.

This record feeds the model card and release decision (Module 11), the EU AI Act's technical documentation for high-risk systems, and any later investigation.

Key terms

Term	Meaning
Use case	A precise statement of a system's purpose, decision, users, affected people and success measures.
Intended purpose	The use a system is designed for, which drives its risk tier and documentation.
AI impact assessment	Assessment of a system's effects on individuals, groups, society and the organization.
Risk register	A log of identified risks with owners, ratings and treatments.
Decision log	A record of design trade-offs, who decided and why.
Provenance	Where data came from and on what terms.
Lineage	How data was transformed and which version trained which model.
Datasheet for datasets	Standard documentation of a dataset's composition, collection, uses and limits.
Acceptance criteria	Thresholds set before testing that a system must meet to be released.
Equalized odds	Fairness measure requiring equal true- and false-positive rates across groups.

Key takeaways

- Start with a precise use case, and ask whether AI is the right tool at all.
- Impact assessment looks at effects on people; coordinate it with DPIAs and other legal assessments.
- Translate each principle into a testable design requirement; log and treat design risks.
- Document as you go: decisions, trade-offs and who made them.
- Data must be lawful, fit, representative and examined for bias; provenance and lineage make it traceable.
- Set acceptance criteria before testing, test by subgroup, and choose fairness measures deliberately.

Knowledge check

12 questions. Choose the single best answer for each.

1. A team proposes a machine-learning model to flag invoices over a fixed approval limit. What is the best governance response at the use-case stage?

- **A.** Approve it, since invoice checks are low-risk
- **B.** Ask whether a simple rule would do the job instead
- **C.** Require a fundamental rights impact assessment
- **D.** Send it straight to red-team testing before any launch

2. A use-case form asks just what the system predicts. Which missing element most limits the ability to assess its risk?

- **A.** The decision it informs and who is affected
- **B.** The programming language and libraries it will use
- **C.** The name of the data scientist leading it
- **D.** The cloud provider that will host it

3. Which question belongs in an AI impact assessment rather than a purely organizational risk assessment?

- **A.** How much will the project cost to build?
- **B.** Could the vendor go out of business during the project?
- **C.** Will customers adopt the product?
- **D.** Could the system deny people access to services?

4. A deployer must run a DPIA, a fundamental rights impact assessment and its own AI impact assessment for one system. What is the best approach?

- **A.** Run each separately with different teams
- **B.** Run one coordinated process meeting each requirement
- **C.** Run the DPIA, since privacy law came first
- **D.** Run the cheapest assessment and cite it for all of them

5. A company's AI policy commits to fairness. Which design requirement best puts that commitment into practice?

- **A.** A statement that the model will be fair
- **B.** A pledge that data scientists will act ethically
- **C.** Named groups to test and a disparity threshold
- **D.** A plan to review fairness after complaints

6. A team finds that a feature it planned to use, postcode, closely tracks applicants' ethnicity. Which risk response changes the design?

- **A.** Accepting the risk with sign-off
- **B.** Transferring the risk to an insurer
- **C.** Avoiding the risk by removing the feature
- **D.** Mitigating it with a notice sent to every customer

7. Two years after launch, a regulator asks why a lender chose a complex model despite needing to explain decisions. What best supports the lender's answer?

- **A.** A decision log made at the time of the choice
- **B.** The current model's accuracy dashboard
- **C.** Interviews with the original developers
- **D.** A fresh impact assessment written for the regulator

8. A model must be retrained after a customer exercises the right to erasure. What makes it possible to find which datasets held the customer's data?

- **A.** The model card's limitations section
- **B.** The acceptance criteria in the test plan
- **C.** Data lineage and provenance records
- **D.** The organization's acceptable-use policy

9. Which data governance requirement addresses a training set in which older applicants are badly under-represented?

- **A.** Representativeness
- **B.** Minimization
- **C.** Retention
- **D.** Confidentiality and access control

10. Why should acceptance criteria be set before testing begins?

- **A.** To shorten the overall testing schedule for the team
- **B.** To let the vendor approve the results
- **C.** To satisfy the cloud provider's terms
- **D.** To stop results being rationalized afterward

11. A model's scores are well calibrated for every group, but its false-positive rate is much higher for one group. What does this illustrate?

- **A.** That the test data was leaked into training
- **B.** That fairness measures can conflict
- **C.** That the model is overfitted to one group
- **D.** That calibration testing is unnecessary

12. After testing, which item belongs in the training and testing record?

- **A.** The marketing plan for the product launch
- **B.** The vendor's customer reference list
- **C.** The board's annual risk appetite statement for AI
- **D.** Subgroup results against acceptance criteria

Answers and explanations

#	Key	Why	Item ID · BoK
1	B	Use-case review includes asking whether AI is the right tool. A fixed threshold is a rule, not a learning problem.	FN-M10-01 · III.A.1
2	A	Risk depends on the decision the output drives and the people affected. Implementation details matter less at this stage.	FN-M10-02 · III.A.1
3	D	Impact assessment focuses on effects on people and society, such as access to services. The others are organizational risks.	FN-M10-03 · III.A.2
4	B	A coordinated process avoids gaps and conflicting conclusions while still satisfying each legal requirement.	FN-M10-04 · III.A.2
5	C	Principles must become concrete, testable requirements: which groups, which measure, what threshold.	FN-M10-05 · III.A.3
6	C	Avoidance means changing the design to remove the source of risk. Notices and insurance do not remove the proxy.	FN-M10-06 · III.A.4
7	A	Contemporaneous records of the trade-off, who decided and why, are the strongest evidence of reasonable care.	FN-M10-07 · III.A.5
8	C	Lineage and provenance trace which data, in which version, trained which model, so affected datasets can be found and removed.	FN-M10-08 · III.B.2
9	A	Representativeness asks whether all relevant groups are present in sufficient numbers. Under-representation leads to weaker performance for that group.	FN-M10-09 · III.B.1
10	D	Fixing thresholds in advance prevents teams from adjusting the bar to fit the results they got.	FN-M10-10 · III.B.3
11	B	When base rates differ, calibration and equal error rates generally cannot both hold. Choosing the measure is a policy decision.	FN-M10-11 · III.B.4
12	D	The record should show results, including by subgroup, against the criteria set in advance, plus issues and how they were resolved.	FN-M10-12 · III.B.5

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 11

Release, Monitoring and AI Incidents

Domain III of the AIGP Body of Knowledge · III.C.1 to III.C.6 · About 50 minutes

This module covers the third competency of Domain III: what a developer does at and after release. It covers the release decision and the documentation that goes with it, monitoring and maintenance, the tests and reviews that probe a live system, how to manage AI incidents and find their root causes, and what to disclose publicly. Many of the same activities reappear for deployers in Module 12.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Describe a release-readiness review, including model cards and, where required, conformity assessment.
2. Plan continuous monitoring, maintenance and a retraining schedule, and explain why retrained models need re-approval.
3. Distinguish audits, red teaming, threat modeling and security testing, and when each is used.
4. Manage and document AI incidents, issues and risks, including legal reporting duties.
5. Find root causes of AI incidents with cross-functional stakeholders.
6. Decide what to disclose publicly to support transparency.

HOW TO USE THIS MODULE

Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. Release readiness (III.C.1)

A release gate is where the organization decides, on the evidence, whether a system may go live. A readiness review typically confirms:

Check	Evidence
Requirements met	Test results against acceptance criteria, including subgroups
Risks treated	Risk register showing each risk mitigated, avoided, transferred or formally accepted
Impact assessment current	Updated assessment reflecting the final design
Documentation complete	Design, data and testing records; instructions for use
Oversight ready	Defined human-oversight roles, trained operators, override and stop procedures
Monitoring ready	Metrics, thresholds, dashboards, alert owners and a retraining schedule
Legal steps done	For example EU AI Act conformity assessment, declaration and registration for high-risk systems
Rollback ready	A way to revert to the previous version or a manual process

Model cards

A **model card** is a short, structured summary for people who will use or rely on a model: its intended use and users, out-of-scope uses, training data summary, evaluation results (including by subgroup), limitations, and ethical considerations. A **system card** does the same for a whole system that combines several models and components. They support transparency to deployers and, when published, to the public.

Release strategies

Risk can be reduced by how a system is released: **shadow deployment** (running alongside the existing process without acting), a **pilot** with limited users, a **staged rollout** to growing shares of traffic, or **champion-challenger** testing against the current model. Each lets problems appear before they affect everyone.

2. Monitoring, maintenance and retraining (III.C.2)

A system's performance can decline after release through **data drift** (inputs change) or **concept drift** (the relationship between inputs and outcomes changes), as Module 1 explained. A monitoring plan, agreed before launch, says:

- **what is measured:** performance, fairness by subgroup, input distributions, override rates, complaints, latency and cost;
- **how often**, and from what data (including ground truth that may arrive late, such as loan defaults);
- **thresholds** that trigger an alert, a review, a rollback or a stop;
- **who acts**, and how fast.

A **retraining schedule** sets when the model is refreshed, whether on a timetable or on triggers such as drift beyond a threshold. Retraining produces a new model version, so it goes through testing and the release gate again; an automatically retrained model that bypasses review is an ungoverned change.

Maintenance also covers updates to code, libraries, third-party models and infrastructure, each of which can change behavior and should pass through change control.

> **CURRENCY NOTE (AS OF NOVEMBER 2026)**
>
> Under the EU AI Act, providers of high-risk systems must run a post-market monitoring system proportionate to the risk (Article 72). The high-risk obligations apply from 2 December 2027 for Annex III systems and 2 August 2028 for Annex I systems (Module 8).

3. Audits, red teaming, threat modeling and security testing (III.C.3)

Activity	What it is	When it fits
Audit	An independent check of whether a system or program meets defined criteria (internal audit, external audit, a legally required bias audit)	Assurance to the board, regulators or customers; periodic or required by law
Red teaming	A deliberate attempt, often by specialists, to make the system fail or misbehave: harmful outputs, jailbreaks, data leakage, misuse	Before release and after major changes, especially for generative and agentic systems
Threat modeling	A structured analysis of how the system could be attacked or abused, and which controls address each threat	At design and when architecture changes
Security testing	Technical testing such as penetration tests, adversarial-example tests and prompt-injection tests	Before release and on a regular schedule

Resources that help: **MITRE ATLAS**, a knowledge base of adversary tactics against AI systems, and the **OWASP Top 10 for LLM Applications**, which lists common risks such as prompt injection, sensitive information disclosure and excessive agency.

> **WATCH THE DISTINCTION**
> Threat modeling is analysis on paper of what could happen; red teaming and security testing try it against the real system; an audit checks whether the whole thing meets criteria. A scenario's verbs usually tell you which: "identify" points to threat modeling, "attack" or "try to break" to red teaming, "verify against criteria" to audit.

4. Managing incidents, issues and risks (III.C.4)

An **AI incident** is an event in which an AI system's development, use or malfunction causes or contributes to harm. A **near miss** is one that could have caused harm but did not. Both should be recorded.

Stage	What happens
Detect	Monitoring alerts, user reports, complaints, audits, press or regulator contact
Triage	Assess severity and scope; assign an owner; decide escalation
Contain	Stop the harm spreading: roll back, switch to a manual process, add human review, restrict use, or shut down

Stage	What happens
Investigate	Establish what happened and why (root cause analysis, section 5)
Remediate	Fix the cause; correct affected decisions; compensate where appropriate
Notify	Meet legal reporting duties and inform affected people, partners and regulators
Learn	Update the risk register, controls, training and documentation

Reporting duties

Several regimes can require reporting: personal data breaches under the GDPR (72 hours to the supervisory authority, Module 6), sector regulators, contracts, and the EU AI Act. Under the AI Act, providers of high-risk systems must report **serious incidents** to the market surveillance authorities of the member states where they occurred (Article 73). The report is due immediately once the provider establishes a causal link, or a reasonable likelihood of one, and in any case no later than **15 days** after the provider (or, where applicable, the deployer) becomes aware of the incident. The limit is **10 days** where a person has died, and **2 days** for a widespread infringement or a serious and irreversible disruption of critical infrastructure.

Issues and risks that are not incidents, such as a test gap discovered later or a new legal requirement, go into the same register so that nothing depends on memory.

5. Root causes, with cross-functional stakeholders (III.C.5)

AI incidents rarely have a single technical cause. Root-cause analysis looks across the whole socio-technical system (Module 1):

Possible cause	Example
Data	Training data unrepresentative of the people the system now serves
Model	Overfitting, or a model update that changed behavior
Context of use	A system applied to a purpose or population it was not designed for
Human oversight	Operators approving outputs without review (automation bias)
Process and governance	A change that skipped review; unclear ownership; missing monitoring
Third parties	A vendor's model change or a compromised component
Attack	Prompt injection, data poisoning or misuse

Techniques such as **five whys** and **fishbone (cause-and-effect) diagrams** help. Because causes span disciplines, the investigation should include data science, engineering, the

business owner, legal, privacy, security, risk and the people who operate the system, and often the affected users' perspective. A blameless approach, focused on systems rather than individuals, gets more honest information.

> **EXAM TIP**
> When a question asks who should investigate an AI incident, options naming one function alone ("the data-science team") are usually wrong. Look for the cross-functional answer.

6. Public disclosures for transparency (III.C.6)

Transparency to the outside world builds trust and is sometimes legally required. Disclosures a developer may make include:

- **model and system cards** describing capabilities and limits;
- **transparency reports** on how systems are governed, tested and monitored;
- **AI use notices** telling people when they are interacting with AI or are subject to AI-assisted decisions;
- **incident disclosures** where people were affected;
- **legally required publications**, such as registration of high-risk systems in the EU database and the training-content summary for general-purpose AI models (Module 8).

Disclosure has limits: it must not expose security weaknesses, trade secrets or personal data, and it must be accurate, since overstated claims are a consumer-protection risk (Module 7). Decide in advance who approves public statements about AI.

Key terms

Term	Meaning
Release gate	The decision point where evidence is reviewed before a system goes live.
Model card	Structured summary of a model's purpose, performance, limits and ethical considerations.
Shadow deployment	Running a system alongside the current process without letting it act.
Retraining schedule	The plan for when and how a model is refreshed, with re-approval.
Red teaming	Deliberately trying to make a system fail or misbehave.
Threat modeling	Structured analysis of how a system could be attacked and how to defend it.
MITRE ATLAS	Knowledge base of adversary tactics and techniques against AI systems.
AI incident	An event in which an AI system causes or contributes to harm.
Serious incident (EU AI Act)	An incident a high-risk provider must report, generally within 15 days.
Root cause analysis	Investigation of the underlying causes of an incident, across data, model, context, people and process.

Key takeaways

- Release through a gate: tested criteria, treated risks, ready oversight, monitoring and rollback.
- Model cards tell users what a model is for and where it falls short.
- Monitor for drift and fairness; retrained models go back through the gate.
- Threat modeling analyzes, red teaming attacks, audits verify against criteria.
- Incidents: detect, triage, contain, investigate, remediate, notify, learn; know the reporting clocks.
- Root causes span data, model, context, people and process, so investigations must be cross-functional.

Knowledge check

12 questions. Choose the single best answer for each.

1. A lender runs a new credit model alongside its existing process for three months, comparing results without letting the model decide. What is this release strategy?

- **A.** Champion-challenger testing
- **B.** A staged rollout
- **C.** Shadow deployment
- **D.** Red teaming

2. Which item belongs in a model card?

- **A.** Uses that are out of scope for the model
- **B.** The salaries of the development team
- **C.** The model's full source code
- **D.** The company's revenue forecast for next year

3. A fraud model is set to retrain itself automatically every week on new data, with no review. What is the main governance problem?

- **A.** Weekly retraining costs too much
- **B.** Each retrained version is an unreviewed change
- **C.** Retraining makes drift impossible to detect
- **D.** Fraud models should not be retrained once in production

4. A credit model's approval rates for one region fall sharply after a local plant closes. What is this most likely to be?

- **A.** A data poisoning attack
- **B.** Prompt injection
- **C.** Overfitting at training
- **D.** Data or concept drift

5. A team lists, in a workshop, the ways an attacker could manipulate its document-reading assistant and the control for each. What is this activity?

- **A.** Threat modeling
- **B.** Red teaming
- **C.** An external audit
- **D.** A penetration test

6. Specialists spend two weeks trying to make a chatbot reveal confidential data and produce harmful advice. What is this?

- **A.** A bias audit
- **B.** Threat modeling
- **C.** Red teaming
- **D.** A model card review

7. An AI scheduling tool double-books operating theaters, but staff catch it before any patient is affected. How should this be handled?

- **A.** Ignore it, since nobody was harmed
- **B.** Record it as a near miss and investigate
- **C.** Report it as a serious incident within 2 days
- **D.** Retrain the model on more data immediately

8. Under the EU AI Act, what is the general deadline for a high-risk provider to report a serious incident?

- **A.** Within 72 hours of the incident
- **B.** Within 30 days of the incident
- **C.** Within 6 months, in its annual report
- **D.** Within 15 days of becoming aware

9. An insurer's claims model wrongly declined claims for weeks. Who should investigate the root cause?

- **A.** The data-science team alone
- **B.** The vendor's support desk alone
- **C.** A cross-functional group including operators
- **D.** Internal audit alone, some months after the fact

10. Investigators find a hiring model performed as tested, but recruiters applied it to a job family it was not trained for. Where does the root cause lie?

- **A.** In the context of use
- **B.** In the training algorithm
- **C.** In the test data split
- **D.** In the cloud hosting

11. Which public disclosure supports transparency without creating new risk?

- **A.** Publishing the full list of known security weaknesses
- **B.** Releasing the training data, including personal data
- **C.** Claiming the system is free of all bias
- **D.** Publishing a model card on capabilities and limits

12. A company wants to publish a transparency report on its AI. What should it decide in advance?

- **A.** Which competitors to compare itself with
- **B.** Who approves public statements about AI
- **C.** How to avoid mentioning any limitations
- **D.** Whether to omit incidents from the report

Answers and explanations

#	Key	Why	Item ID · BoK
1	C	Shadow deployment runs the system in parallel without letting it act, so problems surface before anyone is affected.	FN-M11-01 · III.C.1
2	A	Model cards describe intended and out-of-scope uses, evaluation results, limitations and ethical considerations.	FN-M11-02 · III.C.1
3	B	Retraining creates a new model version, which should be tested and approved through the release gate.	FN-M11-03 · III.C.2
4	D	Changes in real-world conditions after release shift the inputs or their relationship to outcomes. Monitoring should catch this.	FN-M11-04 · III.C.2
5	A	Threat modeling identifies threats and controls systematically. Red teaming and penetration tests attack the real system.	FN-M11-05 · III.C.3
6	C	Red teaming deliberately attacks or misuses a system to uncover harmful behavior before or after release.	FN-M11-06 · III.C.3
7	B	Near misses should be recorded and investigated; they reveal weaknesses before harm occurs.	FN-M11-07 · III.C.4
8	D	The general deadline is 15 days after becoming aware, shortened to 10 days for a death and 2 days for widespread infringements or critical-infrastructure disruption. 72 hours is the GDPR breach clock.	FN-M11-08 · III.C.4
9	C	Causes can lie in data, model, context, oversight or process, so the investigation needs every function that touches the system.	FN-M11-09 · III.C.5
10	A	The model was used outside its intended purpose. Root causes often lie in context or process rather than the model.	FN-M11-10 · III.C.5
11	D	Model cards inform without exposing vulnerabilities or personal data. Overstated claims are a consumer-protection risk.	FN-M11-11 · III.C.6
12	B	Public statements about AI should have a defined approval route, so they are accurate and do not expose sensitive information.	FN-M11-12 · III.C.6

AI LINE OF SIGHT · AIGP FOUNDATIONS

Module 12

Governing AI Deployment and Use

Domain IV of the AIGP Body of Knowledge · IV.A.1 to IV.C.7 · About 60 minutes

Domain IV of the AIGP Body of Knowledge asks how an organization governs AI it deploys, whether bought, licensed, built on someone else's model or built in-house. Most organizations are deployers far more often than developers, so this domain matters to nearly everyone. This module covers the three competencies in turn: the deployment decision, assessing the selected system, and governing deployment and use through to deactivation.

Independent study material. Not affiliated with or endorsed by the IAPP. AIGP is a certification of the IAPP.

Learning objectives

By the end of this module you should be able to:

1. Analyze the context of a use case before choosing how to deploy AI.
2. Compare model types and deployment options, including fine-tuning, retrieval-augmented generation and agentic architectures.
3. Assess the selected system's impact in its own setting, and the risks in vendor and licensing terms.
4. Weigh the risks and opportunities of deploying a proprietary model.
5. Govern deployment and use: policies, monitoring, testing in context, incidents and post-market cooperation.
6. Manage secondary uses, external communication, and deactivation or localization of a system.

HOW TO USE THIS MODULE

Read each section in order, pausing at the exam tips. The key terms and takeaways at the end make a quick revision sheet. Then try the knowledge check without looking back; answers and explanations are on the last page.

1. Context of the use case (IV.A.1)

A deployer cannot govern a system without understanding where and how it will be used. Context questions include:

Question	Why it matters
Who will use it, and with what skill and training?	Determines the oversight and training needed
Who will be affected, and are any of them vulnerable?	Determines the impact and the legal regime
What decision does it inform, and can that decision be reversed?	Determines how much human review is needed
Which laws and sector rules apply in this setting?	A tool lawful in marketing may be high-risk in hiring
What data will it receive, from where?	Determines privacy, security and quality controls
What happens if it fails or is unavailable?	Determines fallback and continuity planning

EXAM TIP

The same product can be low-risk in one deployment and high-risk in another. When a scenario describes a familiar tool in a new setting (a chatbot moved from customer service to medical triage), the context has changed the risk.

2. Model types and deployment options (IV.A.2, IV.A.3)

Model types

Choice	Trade-off
Traditional ML vs. generative	Narrow predictive models are easier to test, bound and explain; generative models are flexible but can hallucinate and are harder to bound
Closed vs. open-weight	Closed models are managed by the vendor but opaque and changeable by it; open-weight models can be inspected, self-hosted and frozen, but the deployer takes on security, safety tuning and maintenance
General-purpose vs. specialized	General models do many tasks adequately; specialized or fine-tuned models do one task better with less unexpected behavior
Large vs. small	Larger models are more capable but cost more, are slower and harder to host privately; small models can run on-premises or on devices
Single-mode vs. multimodal	Multimodal models handle text, images and audio together, which widens both uses and risks

Hosting options

Option	Governance consideration
Vendor API or SaaS	Fast and managed; data leaves the organization; the vendor controls model changes and terms
Private cloud or on-premises	More control over data and versions; the organization carries security, scaling and updates
Edge (on device)	Data stays local and latency is low; harder to monitor, update and recall

Ways to adapt a model

Approach	What it does	Governance consideration
Prompt engineering and system prompts	Shapes behavior with instructions	Cheap and reversible, but can be bypassed; prompts are configuration that needs change control
Retrieval-augmented generation (RAG)	Retrieves relevant documents and supplies them to the model at query time	Grounds answers in approved sources; quality depends on the document store, which needs curation, versioning and access controls that match users' permissions
Fine-tuning	Further trains the model on the organization's data	Better fit for a task; creates a new model to test and document; uses data in training (privacy and IP); may change the organization's legal role (Module 3)
Agentic architecture	Lets the system plan multi-step tasks and act through tools and other systems	Errors become actions: grant least-privilege tool access, require human approval for consequential steps, set spending and action limits, log every action, and keep a way to stop it

WATCH THE DISTINCTION

RAG does not change the model; it changes what the model sees. Fine-tuning changes the model. If a question asks how to keep answers current with policy documents that change weekly, RAG is usually the better fit; if it asks how to teach a consistent style or specialist task, fine-tuning may be.

3. Assessing the selected system (IV.B.1, IV.B.2)

Impact assessment of the selected system

Once a system is chosen, the deployer assesses it **in its own context**: its users, affected people, data, decisions and legal setting. The vendor's documentation and testing are inputs, not substitutes; a vendor cannot know how a particular deployer will configure, integrate and use the system. The assessment should include the deployer's own validation on its own data and population, and should be coordinated with any DPIA or fundamental rights impact assessment required (Modules 6 and 8).

Vendor and licensing terms

Module 5 covered due diligence and contract clauses. Licensing adds model-specific points:

Licence point	What to check
Permitted use	Some licences restrict commercial use, use above a user threshold, or use in particular sectors or for particular purposes
Acceptable-use policy	Prohibited uses incorporated into the licence or terms of service
Attribution and notices	Requirements to credit the model or pass licence terms on to users
Derivatives	Whether fine-tuned versions may be distributed, and under what terms
Outputs	Ownership of outputs and restrictions on using them, for example to train competing models
Data use	Whether the provider may use inputs and outputs, and how long it retains them
Change and termination	How the provider may change or withdraw the model, and with what notice

> **EXAM TIP**
> "Open source" or "open weight" does not mean "no conditions". Many openly available models carry licences with use restrictions. Check the licence before assuming a use is permitted.

4. Proprietary models: risks and opportunities (IV.B.3)

"Proprietary model" is used in two senses, and the exam can use either: a **vendor's closed model** that the deployer licenses, or a model the **organization owns** because it built or fine-tuned it.

	Opportunities	Risks
Vendor's closed model	State-of-the-art capability; the vendor handles hosting, safety tuning and updates; fast to adopt	Opacity; lock-in; vendor changes outside the deployer's control; data leaving the organization; dependence on vendor terms
Organization's own model	Control over data, versions and behavior; differentiation; protection of sensitive data; independence from vendor changes	Full developer burden (data, testing, documentation, security, monitoring); cost and skills; potential provider obligations under AI laws

The choice is a governance decision, not just a technical one. Record the reasoning, who decided, and how the chosen option's specific risks will be managed.

5. Governing deployment and use (IV.C.1 to IV.C.4)

Applying policies and ethics at deployment

- Configure the system to the limits set in the impact assessment (approved uses, data sources, user groups).
- Put human oversight in place with trained operators who have the time, information and authority to intervene.
- Tell users and affected people that AI is involved, as policy and law require.
- Apply acceptable-use rules and access controls.

Monitoring, maintenance and retraining (deployer)

The deployer monitors performance on its own cases, drift, overrides, complaints and fairness by subgroup. It applies vendor updates through change control and re-tests after them. If it fine-tunes or retrains, the new version goes through its own approval process.

Audits, red teaming and security testing (deployer)

The deployer tests what the vendor never saw: its configuration, prompts, integrations, data connections and agent permissions. Prompt injection through the deployer's own documents, data leakage through its own connectors and excessive agency through its own tool grants are deployment risks, not model risks.

Incidents and post-market monitoring

The deployer records incidents, issues and risks in its own register, contains harm, investigates with the provider, and reports where law or contract requires. Under the EU AI Act, deployers of high-risk systems must monitor operation, inform the provider of risks and serious incidents, and suspend use where a system presents a risk; their information feeds the provider's post-market monitoring (Module 11).

6. Secondary uses and downstream harms (IV.C.5)

Systems drift into uses they were never assessed for. A meeting summarizer becomes a performance-evaluation tool; a fraud score is reused to set credit limits; a customer chatbot starts answering medical questions. This is **function creep**, and each new use needs its own assessment.

Downstream harms arise when outputs travel beyond their original use: generated text published as fact, scores shared with third parties, or synthetic content reused out of context. Controls include:

- a clear statement of approved uses, with new uses routed through intake;

- technical limits on who can use the system and for what;
- monitoring of actual use, such as query topics and user groups;
- labeling of AI outputs and restrictions on onward sharing;
- periodic review of whether use has drifted from approval.

7. External communication, deactivation and localization (IV.C.6, IV.C.7)

External communication plans

Decide before go-live how the organization will talk about the system: what customers and affected people are told, how they can ask questions or contest decisions, what is said to regulators and the press, and who speaks for the organization if something goes wrong. An incident is the wrong time to draft the plan.

Deactivation and localization

Deployers need policies and controls to **deactivate** a system, switching it off entirely, or to **localize** it, restricting it to particular regions, user groups, use cases or functions, when a trigger is met. Triggers include:

- performance or fairness falling below thresholds;
- a serious incident or credible risk of harm;
- a legal or regulatory requirement, such as a prohibition in one jurisdiction or a regulator's order;
- a vendor change, contract end or security compromise.

Controls include a tested way to switch off or restrict the system quickly (a "kill switch" or feature flag), a manual fallback process, named people authorized to act, communication to users, and decisions about the system's data and logs. Localization lets an organization comply with a law in one jurisdiction, or contain a problem to one group, without withdrawing a system that works elsewhere.

> **EXAM TIP**
> If a system breaches a new law in one country but is lawful and working well elsewhere, localization (restricting where it is used) is often a better answer than global shutdown.

Key terms

Term	Meaning
Context of use	The users, affected people, decisions, setting and laws around a deployment.
Open-weight model	A model whose trained parameters are published so others can run and modify it.
RAG	Retrieval-augmented generation: supplying relevant documents to a model at query time.
Fine-tuning	Further training a pre-trained model on specific data.
Agentic architecture	A system that plans and acts through tools with limited step-by-step direction.
Least privilege	Granting a system only the access it needs for its task.
Function creep	Gradual use of a system for purposes it was not assessed for.
Downstream harm	Harm arising when outputs are used beyond their original purpose.
Deactivation	Switching a system off, with a fallback in place.
Localization	Restricting a system to particular regions, users, uses or functions.

Key takeaways

- Context decides risk: the same tool can be low-risk in one setting and high-risk in another.
- Model type, hosting and adaptation method each shift who controls risk.
- RAG changes what the model sees; fine-tuning changes the model; agents turn errors into actions.
- Assess the selected system in your own context, and read the licence as closely as the contract.
- "Proprietary model" has two senses; each brings different risks and opportunities.
- Deployers test what vendors never saw: configuration, integrations, prompts and permissions.
- Route new uses through intake, plan communication in advance, and be able to deactivate or localize quickly.

Knowledge check

12 questions. Choose the single best answer for each.

1. A retailer moves its customer-service chatbot onto its pharmacy pages, where customers ask about medicines. What has changed most?

- **A.** The model's underlying architecture
- **B.** The context, and so the risk
- **C.** The vendor's licence terms
- **D.** The chatbot's response speed

2. A bank wants a model it can inspect, host privately and freeze at a known version. Which model type best fits?

- **A.** A closed model accessed through an API
- **B.** A vendor's SaaS product with monthly updates
- **C.** An open-weight model hosted by the bank
- **D.** A consumer chatbot used through a browser

3. An HR assistant must answer questions from a policy handbook that changes every month. Which approach keeps answers current most efficiently?

- **A.** Retrieval-augmented generation over the handbook
- **B.** Fine-tuning the model on the handbook monthly
- **C.** Writing a longer system prompt each month
- **D.** Switching to a larger general-purpose model each month

4. A company gives an AI agent access to its finance system to pay supplier invoices. Which control best limits the risk?

- **A.** Training the agent on more historical invoice data
- **B.** Publishing a model card for the agent
- **C.** Letting the vendor monitor the agent
- **D.** Least-privilege access and approval limits

5. A vendor's documentation says its triage tool performed well in testing. What should the hospital deploying it do?

- **A.** Rely on the vendor's testing, since it is documented
- **B.** Validate the tool on its own patients and setting
- **C.** Ask the vendor to repeat its testing at no cost
- **D.** Defer any testing until patients complain

6. A start-up plans to build its product on an openly available model. What should it check before launch?

- **A.** The licence's permitted-use terms
- **B.** The model's ranking on a public leaderboard
- **C.** Whether the model is the largest available
- **D.** Whether competitors use the same model

7. A bank weighs building its own fine-tuned model against licensing a vendor's closed model. What is the main risk of the closed model?

- **A.** The bank must carry the full developer burden
- **B.** The bank cannot use the model for any decisions about customers
- **C.** The vendor can change the model outside its control
- **D.** The model cannot be accessed over the internet

8. A deployer integrates a vendor's chatbot with its document store and email system. Which risk is most clearly the deployer's to test?

- **A.** The model's general reading comprehension
- **B.** The vendor's training-data provenance records
- **C.** The model's benchmark accuracy scores
- **D.** Prompt injection through its own documents

9. A vendor tells its customers it will update the model behind their contract-review tool next week. What should a deployer do?

- **A.** Re-test the tool through change control
- **B.** Accept the update, since the vendor tested it
- **C.** Ask the vendor to delay updates indefinitely
- **D.** Stop using AI for contract review entirely

10. Managers start using a meeting-summary tool's outputs to rate employees' performance. What is the governance issue?

- **A.** The tool needs a larger language model
- **B.** Function creep into an unassessed use
- **C.** The summaries should be shorter
- **D.** The tool's licence has expired

11. When should a deployer prepare its plan for communicating with customers and regulators about an AI system?

- **A.** After the first complaint arrives
- **B.** During any regulatory investigation
- **C.** Before the system goes live
- **D.** Once the vendor requests it

12. A new law in one country prohibits a feature of an AI system that works well and lawfully elsewhere. What is often the most proportionate response?

- **A.** Shut the system down worldwide
- **B.** Keep using it everywhere until someone challenges it
- **C.** Ask users in that country to consent
- **D.** Localize it by disabling use in that country

Answers and explanations

#	Key	Why	Item ID · BoK
1	B	The same tool in a new setting with health questions and vulnerable users carries different risk and may engage different laws.	FN-M12-01 · IV.A.1
2	C	Open-weight models can be inspected, self-hosted and pinned to a version, though the bank then carries security and maintenance.	FN-M12-02 · IV.A.2
3	A	RAG supplies the current documents at query time, so updating the document store keeps answers current without retraining.	FN-M12-03 · IV.A.3
4	D	Agentic errors become actions. Least-privilege access, approval thresholds and action logs contain the damage.	FN-M12-04 · IV.A.3
5	B	The deployer must assess the selected system in its own context. Vendor testing is an input, not a substitute.	FN-M12-05 · IV.B.1
6	A	Openly available models often carry licences restricting commercial use, certain sectors or derivatives.	FN-M12-06 · IV.B.2
7	C	Closed vendor models bring opacity, lock-in and changes the deployer does not control. The developer burden is the risk of building its own.	FN-M12-07 · IV.B.3
8	D	Integrations, configurations and connected data are deployment risks the vendor never tested. The deployer must red-team its own setup.	FN-M12-08 · IV.C.3
9	A	Vendor updates are changes that can alter behavior. The deployer should re-test on its own cases before relying on the new version.	FN-M12-09 · IV.C.2
10	B	Using a system for a purpose it was not assessed for, here a consequential employment use, is function creep and needs its own review.	FN-M12-10 · IV.C.5
11	C	External communication plans are made in advance, so the organization is not drafting them in the middle of an incident.	FN-M12-11 · IV.C.6
12	D	Localization restricts a system to where it is lawful and safe, avoiding an unnecessary global shutdown.	FN-M12-12 · IV.C.7