

Module 3

Privacy & Data Protection in AI Design



Co-funded by
the European Union



Ethical Engineer

www.ethicalengineer.eu

CONTENTS

- 01 Frame privacy risk in the system
- 02 Design privacy into the architecture
- 03 Govern, assess and provide rights
- 04 Make and defend a privacy decision



This project has been funded with support from the European Commission. The author is solely responsible for this publication (communication) and the Commission accepts no responsibility for any use may be made of the information contained therein. In compliance of the new GDPR framework, please note that the Partnership will only process your personal data in the sole interest and purpose of the project and without any prejudice to your rights.



Ethical Engineer Course at a Glance

Core content and learner output for each module

Module 01	Module 02	Module 03	Module 04	Module 05	Module 06
Ethics & Transparency	Bias, Fairness & Inclusion	Privacy & Data Protection	Responsibility & Safety	Values-based Design	Sustainability & Impact
Ethical principles and EU values Transparency and explainability AI and robotics case analysis	Bias across the AI lifecycle Fairness goals and evidence Inclusive engineering cases	Privacy by design and GDPR Data governance and PETs Assessment and contestability	Roles and accountability Failure, resilience and handover Logging and incident learning	Ethical design requirements Data, model and interface choices Trade offs and agile checkpoints	M Energy, water and hardware Societal and lifecycle impacts Metrics, governance and redesign

Module Orientation

This module turns privacy principles into architecture, evidence and accountable engineering decisions.

Purpose

Equip engineering students to identify privacy risks and design proportionate technical and organisational controls.

Audience

Higher education students in engineering, computer science, robotics and related disciplines.

Builds On

Modules 1 and 2: ethical principles, transparency, bias analysis and engineering decision records.

[Educator Guidance]

Timing: 3 minutes. Explain the module sequence and learner outputs before teaching begins.

Accessibility: provide the learning outcomes and worksheet in advance.

Module Orientation

This module turns privacy principles into architecture, evidence and accountable engineering decisions.

Learning Sequence

Frame privacy risk, design controls, assess obligations and defend a privacy engineering decision.

Learner Outputs

Data-lifecycle map, privacy architecture brief, DPIA screen and engineering decision record.

[Educator Guidance]

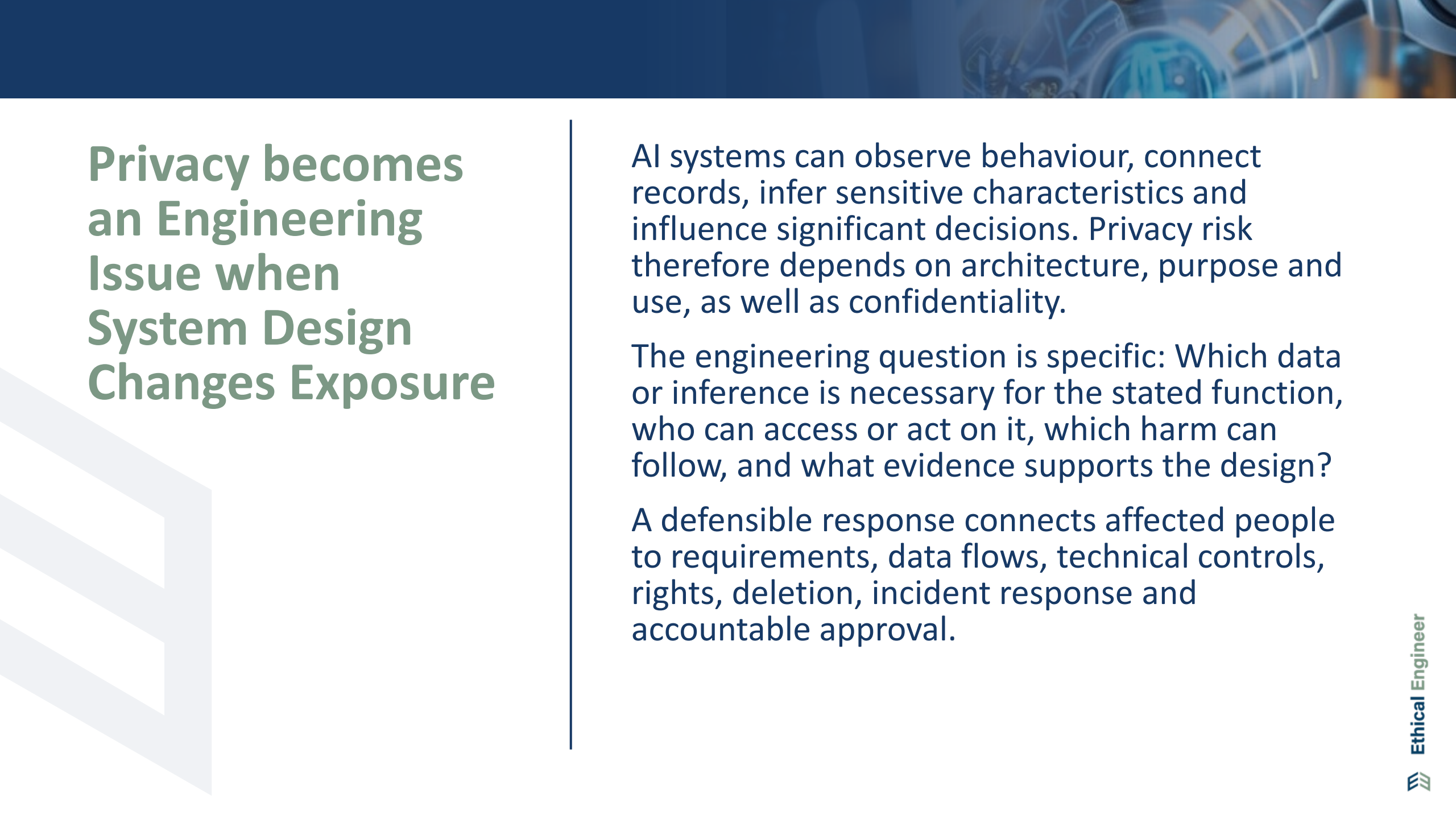
Timing: 3 minutes. Explain the module sequence and learner outputs before teaching begins.

Accessibility: provide the learning outcomes and worksheet in advance.

Learning Outcome

By the end of the module, students will be able to:

1. Map personal data, actors, purposes and privacy risks across an AI system lifecycle.
2. Apply privacy by design, privacy by default and GDPR principles to engineering requirements.
3. Select security and privacy-enhancing technologies for a stated threat and use context.
4. Screen a project for a Data Protection Impact Assessment and design contestability.
5. Document controls, residual risk, monitoring and an accountable engineering decision



Privacy becomes an Engineering Issue when System Design Changes Exposure

AI systems can observe behaviour, connect records, infer sensitive characteristics and influence significant decisions. Privacy risk therefore depends on architecture, purpose and use, as well as confidentiality.

The engineering question is specific: Which data or inference is necessary for the stated function, who can access or act on it, which harm can follow, and what evidence supports the design?

A defensible response connects affected people to requirements, data flows, technical controls, rights, deletion, incident response and accountable approval.

Frameworks Connect Privacy Duties with Operational Evidence

European law and guidance define why privacy must be protected. Operational frameworks help teams turn those duties into repeatable engineering evidence.

EU requirements:

The EU Ethics Guidelines identify privacy and data governance as a condition for trustworthy AI. GDPR establishes principles, rights and accountability. The AI Act adds data governance, documentation and oversight duties for relevant AI systems

Operational support:

NIST AI RMF organises risk work through Govern, Map, Measure and Manage. ISO/IEC 27701:2025 specifies a Privacy Information Management System for PII controllers and processors.

Explain the distinction between requirements and implementation support. European law and guidance establish obligations and expectations. NIST AI RMF and ISO/IEC 27701 can help teams organise risk and privacy-management evidence. Ask learners which artefact in the module demonstrates each requirement.

[Sources]

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
<https://www.nist.gov/itl/ai-risk-management-framework>
<https://www.iso.org/standard/27701>

Frameworks Connect Privacy Duties with Operational Evidence

European law and guidance define why privacy must be protected. Operational frameworks help teams turn those duties into repeatable engineering evidence.

Evidence in this Module:

Data maps, threat models, architecture decisions, DPIA screening, rights routes, tests, owners and review triggers.

Open:

- EU Ethics Guidelines
- GDPR
- EU AI Act
- NIST AI RMF
- ISO/IEC 27701:2025

Explain the distinction between requirements and implementation support. European law and guidance establish obligations and expectations. NIST AI RMF and ISO/IEC 27701 can help teams organise risk and privacy-management evidence. Ask learners which artefact in the module demonstrates each requirement.

[Sources]

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
<https://www.nist.gov/itl/ai-risk-management-framework>
<https://www.iso.org/standard/27701>

01

Frame Privacy Risk in the System

Map purpose, personal data, actors, inferences, sharing, retention and affected people.

Privacy Risk is a System Property

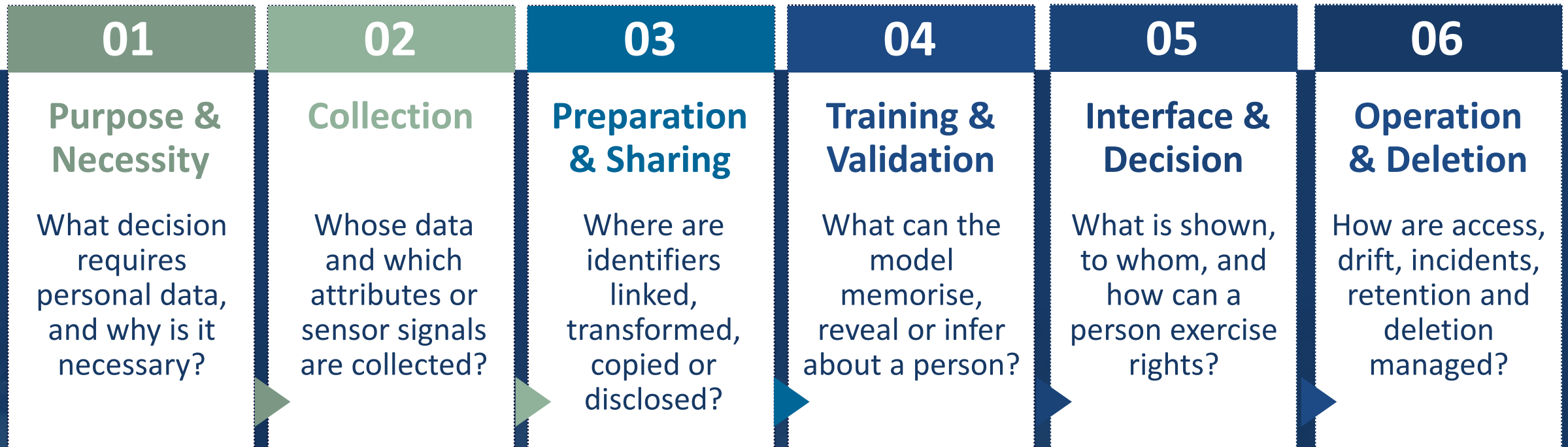
Privacy concerns a person's ability to understand and influence how information about them is collected, connected, inferred, used and retained. A system can be secure against unauthorised access and still create intrusive monitoring, unexpected reuse or an unfair loss of autonomy.

Risk enters through purpose, sensor choice, identifiers, data linkage, model behaviour, interfaces, access rules, vendor relationships and retention.

Engineers therefore examine the complete sociotechnical system, its data flows and the decision it supports.

Map Privacy Across the AI System Lifecycle

Choose one engineering decision. At each stage, identify the data, affected people, privacy risk, evidence and control.



CAPSTONE: A data-lifecycle map linking each stage to purpose, affected people, evidence, privacy risk and a practical control.

Completed Privacy Pathway: Caresecure Fall Detection

Case study:
Ethical Engineer
CareSecure fall detection.

Decision:
Approve in-home monitoring for an older resident. Each stage records data, affected people, privacy risk, evidence and control.

01

Purpose & Necessity

- **Data:** movement needed to detect a fall. Affected: resident and visitors.
- **Risk/Control:** prohibit behavioural surveillance; document the safety purpose and user settings.

02

Collection

- **Data:** room-level movement and a short emergency buffer. Affected: resident, carers and visitors.
- **Evidence/Control:** inventory sensors; process locally and avoid continuous stored video.

[Educator Guidance] Timing: 10 minutes. Reveal one stage at a time and ask learners to predict the next control. Emphasise that the completed example is a reasoned design, not a universal legal answer.

Completed Privacy Pathway: Caresecure Fall Detection

Case study:

Ethical Engineer

CareSecure fall detection.

Decision:

Approve in-home monitoring for an older resident. Each stage records data, affected people, privacy risk, evidence and control.

03

Preparation & Sharing

- **Data:** device ID, alert time and emergency clip. Affected: resident and approved carers.
- **Evidence/Control:** separate identity; encrypt transfer and enforce role-based access.

04

Training & Validation

- **Data:** consented, simulated and representative fall patterns. Risk: memorisation or routine inference.
- **Evidence/Control:** prevent raw export; test leakage, false alerts and operating limits.

[Educator Guidance] Timing: 10 minutes. Reveal one stage at a time and ask learners to predict the next control. Emphasise that the completed example is a reasoned design, not a universal legal answer.

Completed Privacy Pathway: Caresecure Fall Detection

Case study:
Ethical Engineer
CareSecure fall detection.

Decision:
Approve in-home monitoring for an older resident. Each stage records data, affected people, privacy risk, evidence and control.

05

Interface & Decision

- **Data:** alert and confidence. Risk: an alert can reveal health or occupancy.
- **Evidence/Control:** minimise the message; verify before escalation and provide accessible controls.

06

Operation & Deletion

- **Data:** audit log and reviewed incident record. Risk: retention creep, vendor access or breach.
- **Evidence/Control:** automatic deletion schedule, access review, incident route and suspension trigger.

[Educator Guidance] Timing: 10 minutes. Reveal one stage at a time and ask learners to predict the next control. Emphasise that the completed example is a reasoned design, not a universal legal answer.

Completed Privacy Pathway: Caresecure Fall Detection

Case study:
Ethical Engineer
CareSecure fall detection.

Decision

Approve a limited pilot only after a DPIA, explicit operating boundaries, local processing, role-based access, tested alerts, defined deletion and a route to pause or withdraw.

[Educator Guidance] Timing: 10 minutes. Reveal one stage at a time and ask learners to predict the next control. Emphasise that the completed example is a reasoned design, not a universal legal answer.

Personal Data Includes Identifiers, Signals & Reasonable Inferences

- Personal data is information relating to an identified or identifiable person. Names and account numbers are clear examples. Location traces, device identifiers, voice, images, work patterns and combinations of sensor data can also identify or single out a person.
- Special categories under GDPR include health data and biometric data used to uniquely identify a person. AI can create sensitive inferences even when the input appears ordinary.

Classify each field and inference, record why it is needed, and treat identifiability as a context-dependent risk.

Purpose, Minimisation & Lawful Basis Shape the Technical Design

- Purpose limitation requires a specified and legitimate use. Data minimisation limits collection and processing to what is adequate, relevant and necessary for that use. A lawful basis must be identified before processing begins. Consent is one possible basis and may be unsuitable where a person lacks a genuine choice.
- Translate these principles into requirements: remove unused fields, separate optional functions, restrict secondary use, constrain interfaces and set retention and deletion rules.

A useful data field still requires a legitimate purpose, a valid basis, controlled access and evidence of necessity.

Privacy Harm Includes Exposure, Inference, Surveillance & Loss of Agency

- A breach can expose personal information, yet privacy harm can also occur without a breach. Persistent observation can alter behaviour. Data linkage can reveal identity. Models can infer health, beliefs, location or vulnerability. Automated decisions can make a person visible to institutions while the system remains opaque to that person.
- Assess who can learn what, how certain the inference is, which action may follow and whether the person can understand, correct or contest it.

Security protects data. Privacy engineering also limits observation, inference, purpose, power and retention.

Participation Improves Privacy Requirements

- Affected people can reveal privacy expectations that a technical team may not see. Residents, workers, patients, visitors, operators, maintainers and bystanders may experience the same sensor or model differently.
- Use interviews, co-design, representative testing or an advisory panel to examine acceptable purpose, private spaces, visible indicators, controls, accessible notices, withdrawal, correction and remedy. Participation must have a defined route into requirements and approval.

Record whose views changed the design and which unresolved concern remains.

Worked Mechanical Engineering

Example: Privacy-aware Exoskeleton

A factory proposes a sensor-equipped industrial exoskeleton to improve fit, detect unsafe loading and plan maintenance. The privacy design determines whether the system supports workers or becomes a monitoring tool.

[Educator Guidance]

Timing: 12 minutes. Use the example to keep privacy anchored in mechanical engineering. Ask which requirement would be written into the device specification and procurement contract.

01

Initial Brief

Collect joint angles, forces, device status and heart rate so engineers can improve fit and predict component failure.

02

Privacy Risk

Biomechanics can reveal health, disability, fatigue and pace. Workers, contractors and visitors may be identifiable.

03

Reframed Purpose

Support equipment safety and individual fit. Exclude productivity scoring, attendance enforcement and disciplinary use.

Worked Mechanical Engineering

Example: Privacy-aware Exoskeleton

A factory proposes a sensor-equipped industrial exoskeleton to improve fit, detect unsafe loading and plan maintenance. The privacy design determines whether the system supports workers or becomes a monitoring tool.

[Educator Guidance]

Timing: 12 minutes. Use the example to keep privacy anchored in mechanical engineering. Ask which requirement would be written into the device specification and procurement contract.

04

Architecture

Extract safety features on the device, separate worker identity, overwrite raw sensor buffers and encrypt maintenance events.

05

Access & Rights

Give workers their own record. Provide engineers aggregated trends. Restrict individual health interpretation to authorised staff.

06

Decision & Review

Pilot only with documented lawful basis, worker representation, DPIA screen, deletion schedule, complaint route and stop trigger.

Worked Mechanical Engineering Example: Privacy-aware Exoskeleton

A factory proposes a sensor-equipped industrial exoskeleton to improve fit, detect unsafe loading and plan maintenance. The privacy design determines whether the system supports workers or becomes a monitoring tool.

Privacy Changed the Design

A condition-monitoring task became a bounded safety system with local processing, worker rights and a prohibited-use specification.

[Educator Guidance]

Timing: 12 minutes. Use the example to keep privacy anchored in mechanical engineering. Ask which requirement would be written into the device specification and procurement contract.

02

Design Privacy into the Architecture

Apply privacy by design, lifecycle security and suitable privacy-enhancing technologies.

Privacy by Design is an Engineering Method

Privacy by design embeds data protection into the system and its operating model at the earliest design stage. Privacy by default ensures that the protective configuration applies without requiring a person to find and change settings.

The seven privacy-by-design principles provide useful direction. Engineering work turns them into requirements, interfaces, tests, records and controls across the complete lifecycle.

[Educator guidance]

Timing: 8 minutes. Ask groups to connect each of the seven privacy-by-design principles with one engineering artefact.

Privacy by Design is an Engineering Method

01

Purpose

Define the function, affected people, necessary data and prohibited uses. Challenge any collection that lacks a specific need.

02

Architecture

Embed minimisation, separation, local processing, access control, secure defaults, retention and deletion into the system.

03

Accountability

Verify controls, document decisions, provide rights and remedy, monitor use and respond to incidents and change.

Privacy by design is a continuous engineering method. Proactive prevention, default protection, embedded controls, positive-sum design, lifecycle security, transparency and respect for the person must remain visible in evidence and operation.

Privacy by Default Sets the Safe Baseline

- Default settings influence real outcomes because many people never change them. A privacy-protective default collects the minimum data, exposes the smallest audience, enables the least intrusive function and retains information for the shortest justified period.
- Optional analytics, sharing or secondary use should require a clear and separate choice where consent is the appropriate basis. Turning a feature off should stop collection and trigger the stated retention rule.

Test the first-use experience, reset behaviour, software update and administrator configuration. The default must survive normal operation.

A Data-Flow Map Reveals Trust Boundaries & Hidden Copies

- Begin with the physical and digital system. Record each sensor, user input, derived feature, model, log, cache, dashboard, export, vendor interface and backup. For every flow, identify the sender, recipient, purpose, identifier, location, access rule and retention.
- Mark trust boundaries where data moves between a device and cloud, one organisation and another, or an operational system and a research environment. Include failure paths such as offline queues, debug logs and support access.

The map becomes evidence for minimisation, security design, DPIA, contracts, deletion testing and incident response.

Pseudonymisation Reduces Linkage Risk but does not Remove GDPR Duties

- Pseudonymisation replaces direct identifiers and keeps the additional information separately protected. The data remains personal data when re-linking is reasonably possible. It can reduce exposure and support separation of duties.
- Anonymisation aims to prevent identification by means reasonably likely to be used. Removing names alone is inadequate where unique trajectories, rare conditions or linked datasets can identify a person.

Document the attacker, auxiliary data, linkage pathway and re-identification test. Reassess when new datasets or capabilities become available.

Lifecycle Security Supports Confidentiality, Integrity & Controlled Use

- Encryption protects data in transit and at rest, while key management determines whether that protection is real. Role-based access, least privilege, multi-factor authentication, secure update, logging and review reduce inappropriate access. Integrity checks and provenance protect decisions against altered or incomplete data.
- Apply controls to training data, model artefacts, prompts, outputs, logs, backups and deletion queues. Test vendor and administrator access as well as external attacks.

A complete control states the threat, component, configuration, evidence, owner, failure response and review trigger.

Privacy Attack Surfaces Extend Beyond Stored Data

AI can leak personal information through training records, model behaviour, updates, interfaces and metadata. Encryption protects important paths, while other paths require model-specific controls.

Model Inversion or Reconstruction:

Queries, outputs, gradients or parameters may reveal attributes or approximate training examples.

Membership Inference:

An attacker tests whether a specific person or record formed part of the training data.

Gradient Leakage:

Shared updates in distributed learning can reveal information about local examples.

[Sources]

- <https://csrc.nist.gov/pubs/ai/100/2/e2025/final>
- <https://www.enisa.europa.eu/sites/default/files/publications/ENISA%20Report%20-%20Securing%20Machine%20Learning%20Algorithms.pdf>

Privacy Attack Surfaces Extend Beyond Stored Data

AI can leak personal information through training records, model behaviour, updates, interfaces and metadata. Encryption protects important paths, while other paths require model-specific controls.

Scraping and side channels:

Exposed endpoints, logs, packet size, timing or other metadata can reveal data, model behaviour or sensitive topics.

Engineering response:

Define attacker access, restrict outputs and queries, secure updates, test leakage, monitor abuse and state when deployment must stop.

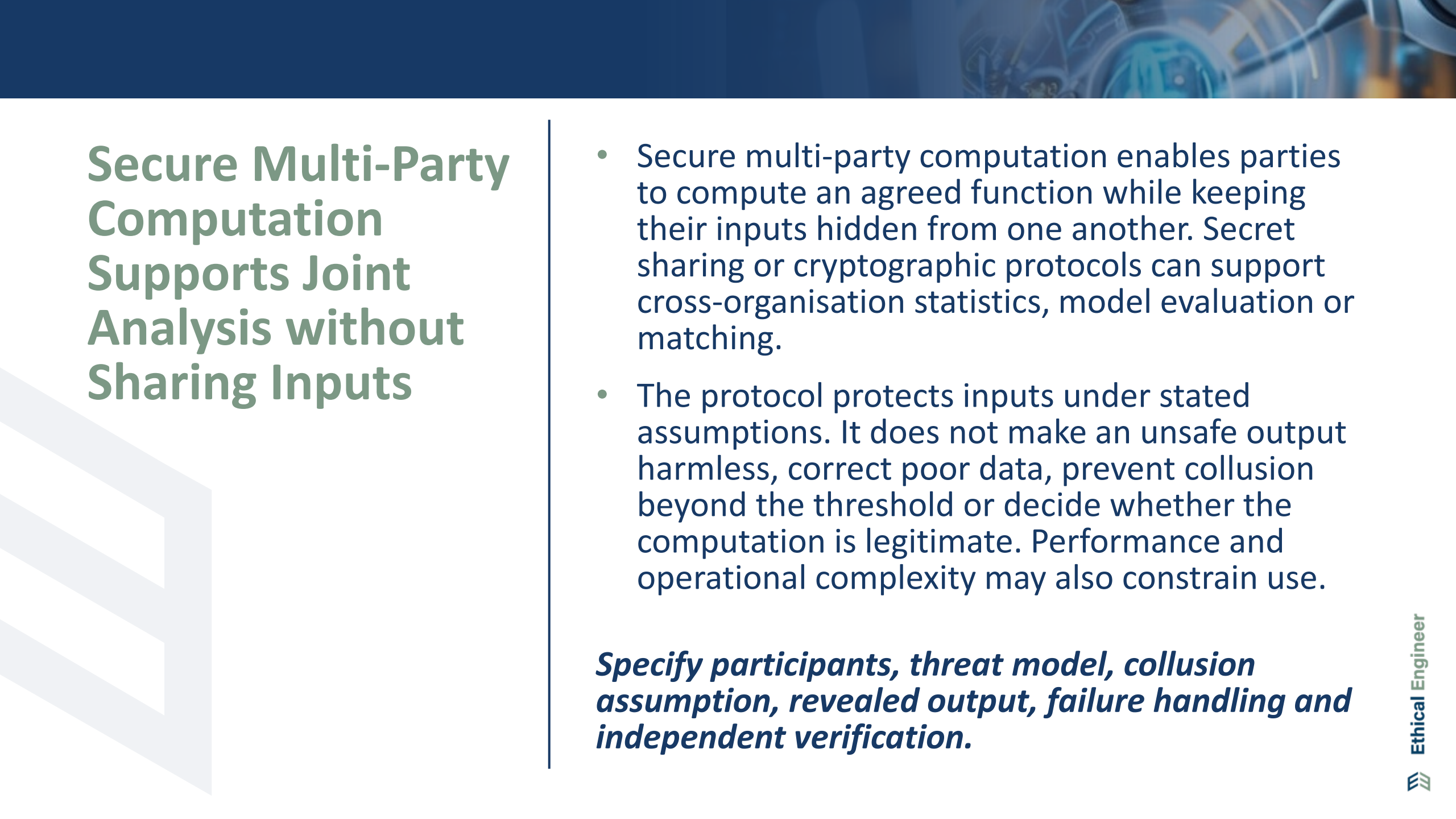
[Sources]

- <https://csrc.nist.gov/pubs/ai/100/2/e2025/final>
- <https://www.enisa.europa.eu/sites/default/files/publications/ENISA%20Report%20-%20Securing%20Machine%20Learning%20Algorithms.pdf>

Federated Learning keeps Raw Training Data Local but Needs Added Safeguards

- Federated learning allows devices or organisations to contribute model updates without pooling their raw datasets. It can reduce central collection and support data localisation.
- Model updates may still reveal information through gradient leakage, membership inference or a malicious participant. Central orchestration, update authentication and device compromise remain part of the threat model.

Use secure aggregation, update clipping, differential privacy, participant authentication and leakage testing where justified. Record what leaves each device and what the coordinator can observe.



Secure Multi-Party Computation Supports Joint Analysis without Sharing Inputs

- Secure multi-party computation enables parties to compute an agreed function while keeping their inputs hidden from one another. Secret sharing or cryptographic protocols can support cross-organisation statistics, model evaluation or matching.
- The protocol protects inputs under stated assumptions. It does not make an unsafe output harmless, correct poor data, prevent collusion beyond the threshold or decide whether the computation is legitimate. Performance and operational complexity may also constrain use.

Specify participants, threat model, collusion assumption, revealed output, failure handling and independent verification.

Differential Privacy Provides a Measurable Bound on Individual Contribution

- Differential privacy adds calibrated randomness so that the presence or absence of one person's data has a limited influence on the released result. The privacy parameter and composition across repeated releases determine the strength of the guarantee.
- Noise affects utility, especially for small groups or rare events. A claimed guarantee is incomplete without the mechanism, unit of protection, privacy budget, clipping rule, composition and implementation evidence.

Use differential privacy for a defined release or learning process, then test both privacy leakage and engineering performance.

Select a Privacy-Enhancing Technology for the Stated Threat

- A privacy-enhancing technology is a design control, not a general label. Select it after defining the data flow, attacker, capability, protected information and required output.
- Federated learning reduces raw-data centralisation. Secure multi-party computation hides inputs during a joint computation. Differential privacy limits information revealed by outputs or updates. Encryption, access control, secure hardware and local processing may be required alongside them.

Compare privacy guarantee, assumptions, performance, failure modes, operational burden and verification. Combine controls only when each addresses a documented risk.

Edge, IoT & Generative AI Change the Privacy Threat Model

01

Edge & IoT Exposure:

Local processing can reduce cloud transfer. Field devices may face physical access, weak identity, resource limits, intermittent patching, distributed logs and difficult deletion

02

Edge & IoT Controls:

Use secure boot, device identity, least data, short local buffers, signed updates, network segmentation and a complete lifecycle inventory.

03

Generative AI Controls:

Define approved data paths, redaction, access separation, retention, provider terms, output review and leakage tests.

[Sources]

- <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence>
- <https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.sp.1800-36.pdf>

Edge, IoT & Generative AI Change the Privacy Threat Model

04

Generative AI Exposure:

Prompts, retrieved documents, fine-tuning data, outputs and logs may contain personal or confidential information. Models may memorise, infer or produce inaccurate statements about people.

05

Decision Check:

Record what stays local, what leaves, what can be inferred, how controls are verified and who can suspend use.

[Sources]

- <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence>
- <https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.sp.1800-36.pdf>

Project Case: Faceboard Airport Facial Recognition

Open the [Ethical Engineer FaceBoard case study](#) and identify which architecture can meet the boarding purpose with the least privacy intrusion.

- The Ethical Engineer FaceBoard case proposes facial recognition for airport boarding. The system may reduce queues and fraud, while processing biometric data in a setting where refusal can affect travel. The case raises purpose, necessity, bias, breach, retention and alternative-route questions.
- **Engineering Focus:** on-device template creation, central storage, one-to-one verification, deletion after travel, vendor access, presentation-attack detection, human review and a non-biometric route.

[Educator guidance] Timing: 12 minutes. Learners should read the linked case before class or use the summary on the slide. Provide a text transcript and allow private reflection before discussion.

Applied Task: Write a Privacy Architecture Decision

[Educator guidance]

Timing: 25 minutes. Assess purpose clarity, data-flow accuracy, threat-control fit, testability, rights and residual-risk reasoning.

- Use FaceBoard or the exoskeleton example. Produce a design brief that another engineering team could implement and test.
- Define the function, affected people, personal data, purpose, lawful-basis question, trust boundaries and threat model. Compare two architectures. Specify minimisation, separation, access, encryption, retention, deletion, rights, monitoring and incident controls.

Add acceptance tests, an accountable owner, residual risk and a condition that requires REVISE or STOP.

Output

One-page architecture brief and a two-minute evidence-based justification.

03

Govern, Assess & Provide Rights

Connect GDPR, DPIA, contestability and AI Act data governance with engineering evidence.

GDPR Principles Become Engineering Requirements

- GDPR Article 5 sets principles for lawful, fair and transparent processing; purpose limitation; data minimisation; accuracy; storage limitation; integrity and confidentiality; and accountability. These principles apply together.
- Translate them into a data specification, allowed-use policy, data-quality test, access model, retention schedule, deletion test, notice, rights route, incident process and decision record. Define controller, processor and supplier responsibilities.

Compliance requires organisational judgement. Engineers contribute evidence about how the architecture behaves in practice.

Lawful Basis, Consent & Transparency Require Coherent Operation

- A controller must identify an appropriate lawful basis for each processing purpose. Consent requires a genuine choice, specific information and an effective withdrawal mechanism. Power imbalance can make consent unsuitable, including some employment and essential-service settings. Special-category data requires an additional condition.
- Transparency should explain what is collected, why, who receives it, how long it is kept, which rights apply and how to exercise them. The interface and actual data flows must match the notice.

Test withdrawal, correction and deletion end to end, including derived data, vendors, backups and model pipelines.

Automated Decisions Require Meaningful Review & Contestability

- GDPR Article 22 concerns certain decisions based solely on automated processing that produce legal or similarly significant effects, subject to stated conditions and exceptions. Related transparency and access provisions can require meaningful information about the logic and consequences in relevant cases.
- A review route needs an authorised and competent person, access to relevant evidence, time to change the outcome, freedom to disagree with the system and a recorded reason. A button that cannot alter the decision is not effective human intervention.

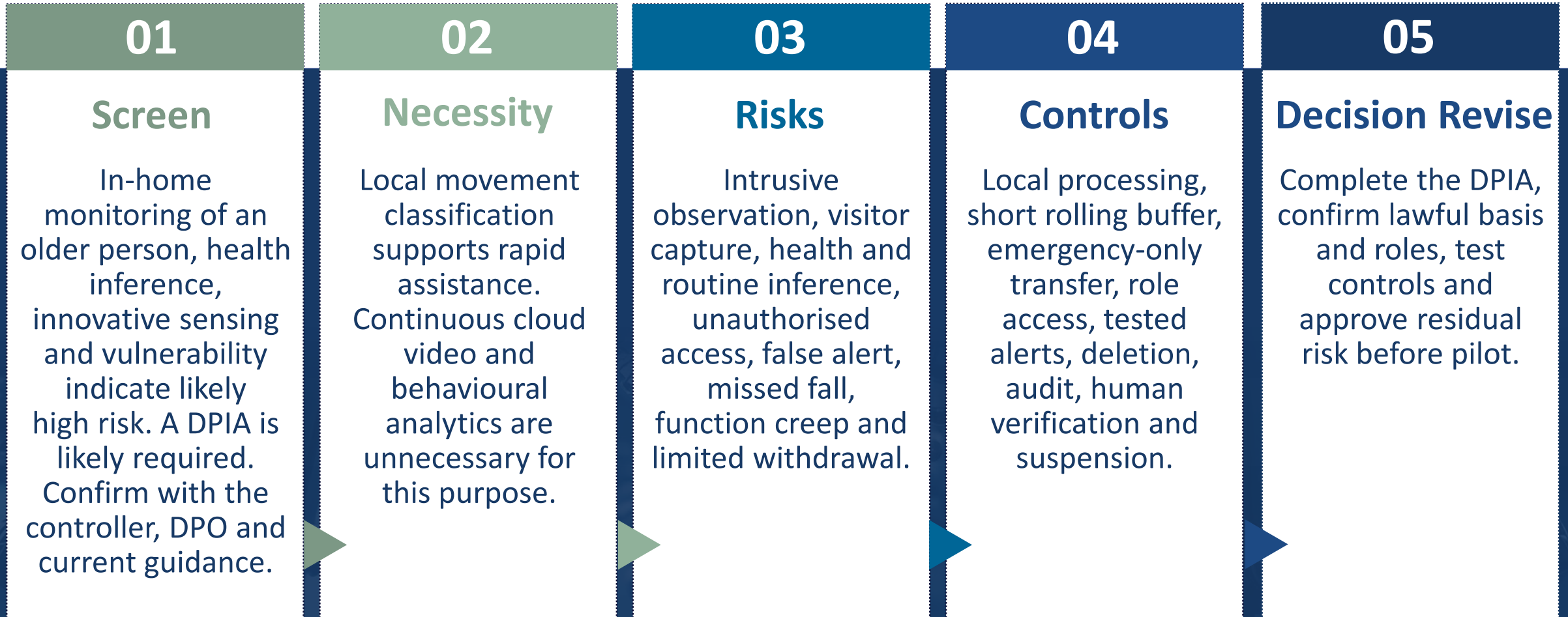
Design notice, explanation, correction, escalation and remedy as system functions, then test them with affected users.

A DPIA Connects High-risk Processing With Design Decisions

- A Data Protection Impact Assessment is required where processing is likely to result in a high risk to people's rights and freedoms. Screening should consider systematic monitoring, innovative technology, vulnerable people, large-scale special-category data, data matching and significant automated decisions.
- A DPIA describes the processing and purpose, assesses necessity and proportionality, identifies risks, records measures and residual risk, and involves relevant expertise and affected people where appropriate.

Begin early enough to change requirements and architecture. Revisit the assessment when purpose, data, model, scale, supplier or context changes.

Completed DPIA Screen: CareSecure In-Home Fall Detection



EU AI Act Data Governance & GDPR Privacy Duties Interact

- For relevant high-risk AI systems, the AI Act requires risk management, data governance, technical documentation, logging, human oversight, accuracy, robustness and cybersecurity. Training, validation and test datasets require appropriate governance concerning relevance, representativeness, errors and gaps.
- GDPR and the AI Act have distinct scope. Data quality does not remove purpose limitation or minimisation. A DPIA and a fundamental-rights impact assessment have different legal triggers and contents, although evidence and controls may be coordinated.

Maintain a traceable control set that records each requirement, evidence item, owner, residual risk and review point.

HealthTech Privacy Evidence is also Patient-safety Evidence

When AI software has a medical purpose, privacy decisions interact with the MDR, the AI Act and GDPR. Intended purpose and risk determine the applicable device obligations.

01 Data Integrity

Trace source, lawful access, labelling, exclusions, transformations, missingness and version. Corrupted or unrepresentative clinical data can change diagnosis, triage or treatment.

02 Quality & Risk Management

Connect data controls to the quality management system, risk-management file, clinical evaluation, cybersecurity, change control and post-market monitoring.

03 Patient Privacy

Restrict special-category data, separate identities, verify access and deletion, and protect logs, model artefacts and supplier interfaces.

HealthTech Privacy Evidence is also Patient-safety Evidence

When AI software has a medical purpose, privacy decisions interact with the MDR, the AI Act and GDPR. Intended purpose and risk determine the applicable device obligations.

04 Evidence before Release

Clinical performance by relevant group, data lineage, threat testing, human review, incident route, residual risk and approval authority.

05 Engineering Conclusion

A privacy failure that undermines data integrity, availability or appropriate use can become a device safety failure.

[Sources]

- <https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng>
- https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en
- <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
- <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

Dilemma: Safety Investigation or Function Creep?

- An exoskeleton records a near miss in which a worker lifts outside the approved range. The safety team requests the previous three months of individual biomechanics to investigate equipment performance. A line manager asks for the same data to identify workers who are slow or non-compliant. The original notice covers fit, safety and maintenance.
- Decide GO, REVISE or STOP for each request. Define purpose, necessity, authorised access, data window, worker involvement, evidence, deletion and review.

Explain which technical separation and governance rule prevents a legitimate safety function becoming performance surveillance.

[Educator guidance] Timing: 15 minutes. Require separate decisions for the safety and performance requests. Invite a dissenting view and record the governance control that resolves the conflict.

04

Make & Defend a Privacy Decision

Compare project cases, complete a design review and defend a decision using evidence.

Project Case A: AI in Healthcare & In-Home Monitoring

- CareSecure uses local sensors to detect a fall and alert a caregiver. The safety benefit is substantial, while continuous observation inside a home can affect privacy, autonomy and trust. Visitors and carers may also be captured.
- Define the minimum signal required, private spaces, local processing, emergency disclosure, consent or other lawful-basis questions, accessible controls, false-alert review, deletion and incident response.

***Open the Ethical Engineer healthcare case study.
Decide which evidence is required before a limited pilot
and which use must be prohibited.***

[Educator guidance] Timing: 12 minutes. Assign resident, caregiver, engineer and data protection roles. Accessibility: allow written role statements instead of live performance.

Project Case B: AI Witness & Sensitive Testimony

- AI Witness records witness statements through a conversational system. The service may preserve details and reduce delay, while collecting highly sensitive accounts that can affect safety and justice. The interview process may also influence what a distressed person reveals.
- Define purpose, data categories, secure capture, language access, human review, correction, legal disclosure, retention, breach impact, model-provider access and a non-automated alternative.

Open the Ethical Engineer witness testimony case study. Decide whether the design should support, constrain or reject automated interviewing.

[Educator guidance] Timing: 12 minutes. Use a trauma-informed approach. Learners may analyse system design without role-playing a distressed witness.

Compare the Cases Before Selecting Privacy Controls

- CareSecure concerns persistent observation in a private home for an immediate safety function. AIWitness concerns a time-limited but highly sensitive interaction whose record may enter a justice process. Both require security, yet their purposes, power relationships, retention, rights and consequences differ.
- For each case, state the data subject, bystanders, controller, purpose, high-risk indicators, trust boundaries, worst credible privacy harm and required human authority.

A useful conclusion may choose different architectures, lawful-basis questions, evidence thresholds and stop conditions. Explain why each control fits the context.

[Educator guidance] Timing: 10 minutes. Compare contexts before controls. Credit well-justified differences rather than a single preferred answer.

Team Activity: Complete a Privacy Design Review

Choose CareSecure, FaceBoard, AIWitness or the exoskeleton. Assign systems, data, security, rights, stakeholder and recorder roles.

1. Define the function, purpose and affected people.
2. Complete the six-stage data-lifecycle map.
3. Draw data flows and trust boundaries.
4. Screen for a DPIA and select proportionate controls.
5. Specify rights, evidence, residual risk and monitoring.
6. Decide GO, REVISE or STOP.

Output:

Decision record and a two-minute privacy brief.

[Educator guidance]

Timing: 30 minutes. Use groups of four to six. Require each role to contribute one piece of evidence or one unresolved concern.

What a Strong Privacy Design Contains

- A strong design begins with a specific function and demonstrates why each personal-data element is necessary. It distinguishes personal data, sensitive inference and non-personal system data. It maps actors, copies, trust boundaries, purposes and retention.
- The architecture combines minimisation, separation, security, privacy-enhancing technology, usable rights, human review, deletion, monitoring and incident response. Each control has a threat, configuration, acceptance test and owner.

The decision record shows lawful-basis and DPIA questions, stakeholder input, alternatives, residual risk, approval authority, review trigger and prohibited uses

Worksheet: Privacy Engineering Decision Record

System, function and affected people:

Specific purpose and prohibited uses:

Personal data, sensitive inferences and necessity:

Data flows, recipients and trust boundaries:

Lawful-basis and DPIA screening questions:

Technical control and acceptance test:

[Educator guidance] Timing: 20 minutes. Provide the worksheet as an editable document.
Accept typed, handwritten, audio or recorded completion.

Continues On Next Slide



Rights, human review, deletion and remedy:

Accountable roles and evidence owner:

Decision: GO REVISE STOP

Residual risk, monitoring and review trigger:

Dissent or alternative architecture:

[Educator guidance] Timing: 20 minutes. Provide the worksheet as an editable document. Accept typed, handwritten, audio or recorded completion.

Role Play: DPIA & Procurement Review Panel

- A university plans to procure an AI camera system that records laboratory activity, detects unsafe behaviour and generates individual compliance scores. The vendor offers cloud processing and retains clips for model improvement. The proposal provides overall accuracy but no evidence about visitors, disability, protective clothing, deletion or staff access.
- Roles: laboratory engineer, student representative, safety lead, data protection officer, vendor, accessibility adviser and procurement lead.

The data flows, DPIA questions, pilot boundary, contract conditions and acceptance evidence. State one condition for REVISE and one for STOP. Record who can suspend use and how a person can access, correct or contest a record.

[Educator guidance] Timing: 25 minutes. The panel must state approval conditions, evidence owners and suspension authority. Do not reward a generic compliance list.

Individual Reflection & Transfer

Write a short engineering argument

1. Which data flow or inference created the strongest privacy concern?
2. Which design choice reduced the risk most effectively?
3. Which lawful-basis or DPIA question remains for specialist review?
4. Which evidence would change your decision?
5. Which use should be prohibited even if technically possible?
6. How would the analysis change in another engineering discipline?

Write:

Submit written, audio or recorded evidence. Connect affected people, engineering action, evidence and a justified decision.

[Educator guidance] Timing: 10 minutes. Learners may submit written, audio or recorded evidence. Give quiet preparation time before any sharing.

Seven Privacy-By-Design Principles as Engineering Checks

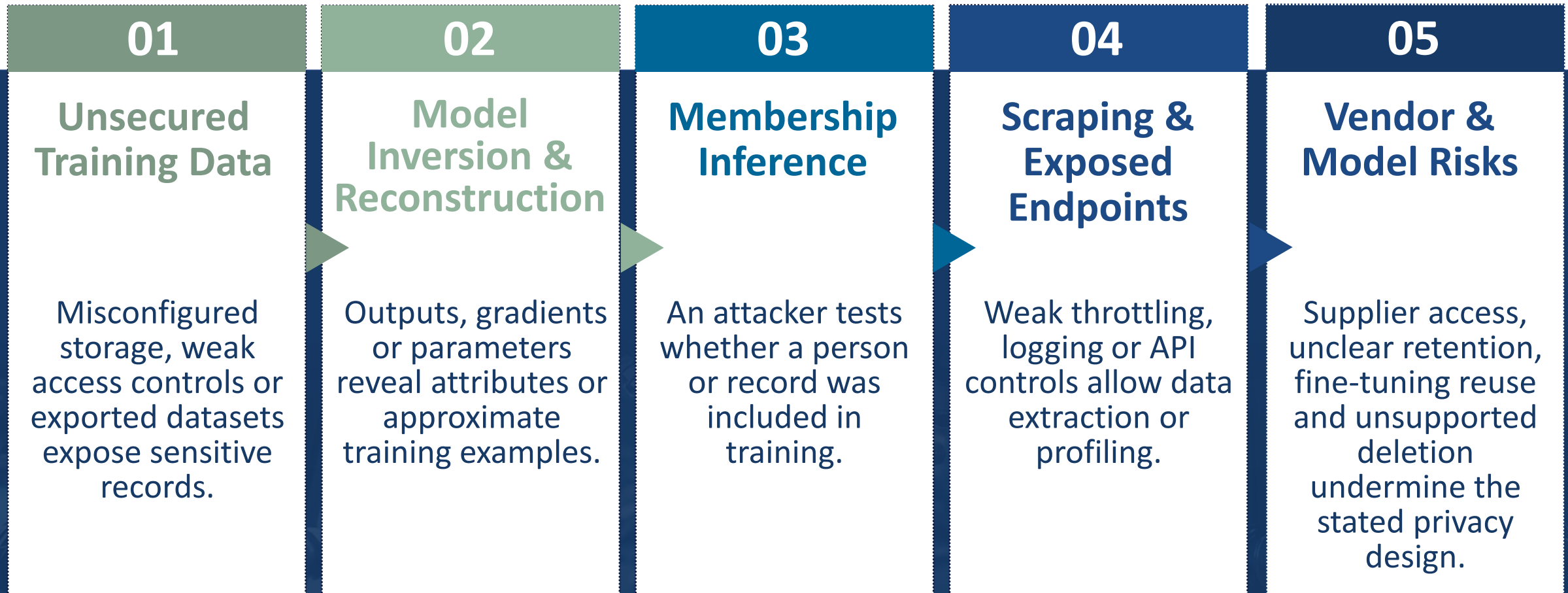
Use the seven principles as a quick diagnostic before approving an AI design.

- **Proactive prevention:** identify likely privacy failures before build or deployment.
- **Default protection:** collect, expose and retain the least data by default.
- **Embedded privacy:** put minimisation, separation, access, deletion and PETs into the architecture.
- **Positive-sum design:** meet the engineering function without treating privacy as a trade-off to remove.
- **End-to-end security:** protect data, model artefacts, prompts, logs, backups and deletion queues.
- **Visibility, transparency and user-centricity:** document controls, explain effects and provide usable rights.

Side-Channel & Metadata Leakage in AI Systems

- Encryption protects content in transit and at rest, but AI systems can still leak information through metadata, timing, query behaviour or output patterns.
- **Example risk:** an LLM conversation protected by TLS may still reveal sensitive topic patterns through packet size or timing analysis.
- **Engineering response:** define metadata as part of the privacy attack surface, reduce exposed logs, restrict query behaviour and monitor abnormal access.
- Where risk is high, test inference from outputs, prompts, gradients, packet patterns and operational telemetry.
- Acceptance evidence should show that leakage testing, monitoring and incident response cover model-specific attack paths.

Operational AI privacy vulnerabilities to test



Governance Triad & Operational Standards

A mature AI privacy process connects legal, privacy and cybersecurity expertise from the start of system design.

01

Legal

Lawful basis, rights, contractual liability, regulatory alignment and documentation obligations.

02

Privacy

Data minimisation, DPIA screening, transparency, rights routes, retention and privacy-management evidence.

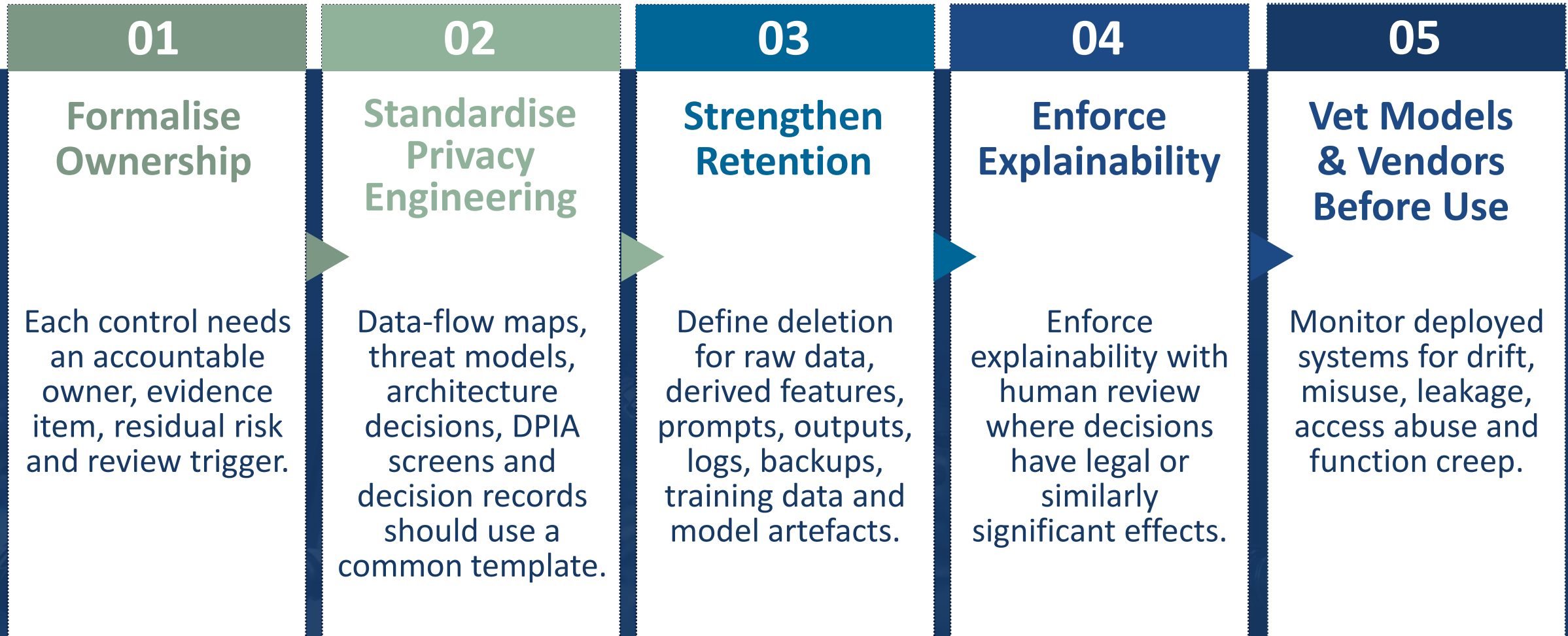
03

Cybersecurity

Confidentiality, integrity, availability, adversarial testing, supplier assurance, monitoring and incident response.

Use NIST AI RMF to organise risk work and ISO/IEC 27701 to structure privacy-management evidence for controllers and processors.

Strategic Implementation Checklist



Optional Ready-To-Use Teaching Activities

01

Teachable Machine:

Explore biometric privacy and consent using image or voice classification.

02

What-If Tool:

Inspect subgroup effects, thresholds and counterfactuals involving sensitive attributes.

03

Facets:

Audit datasets for imbalance, anomalies, missing values And unnecessary fields.

Optional Ready-To-Use Teaching Activities

04

ARX:

Demonstrate pseudonymisation, k-anonymity and re-identification trade-offs.

05

H5P Branching Scenarios:

Simulate consent, surveillance and data-use decisions.

06

DEDA or Ethical OS:

Structure stakeholder dialogue and long-term privacy-risk reflection.

Quiz

1. Why can a system be secure and still create a privacy harm?
2. What is the difference between pseudonymised and anonymised data?
3. Why must a PET be linked to a threat model?
4. What should a data-flow map include?
5. Name two indicators that can support a DPIA screening decision.
6. What makes human review meaningful for an automated decision?
7. What belongs in a privacy engineering decision record?

[Educator guidance]. Timing: 8 minutes. Use individual retrieval first, then pair discussion. The next slide gives concise answers and rationale.

Quiz Answers & Teaching Rationale

1. Secure processing can still enable intrusive observation, inference, reuse or loss of agency.
2. Pseudonymised data remains linkable with additional information; anonymised data should resist identification by reasonably likely means.
3. Each PET protects particular information under stated assumptions and has distinct failure modes.
4. Sources, recipients, purposes, identifiers, transformations, copies, trust boundaries, access and retention.
5. Systematic monitoring, vulnerable people, innovative technology, sensitive data, matching or significant automated decisions.
6. Authority, competence, evidence, time, freedom to disagree and power to change the outcome.
7. Purpose, data, risks, controls, evidence, roles, rights, residual risk, monitoring, decision and review trigger.

Learning Outcome Check

Before finishing, confirm that you can:

1. Map personal data, actors, purposes, inferences and risks across an AI lifecycle.
2. Apply privacy by design, privacy by default and GDPR principles to requirements.
3. Select and justify a privacy-enhancing technology for a stated threat.
4. Screen for a DPIA and design meaningful rights and contestability.
5. Document controls, evidence, residual risk, monitoring and an accountable decision.

Evidence:

Data-lifecycle map, architecture brief, DPIA screen and engineering decision record.

[Educator guidance]

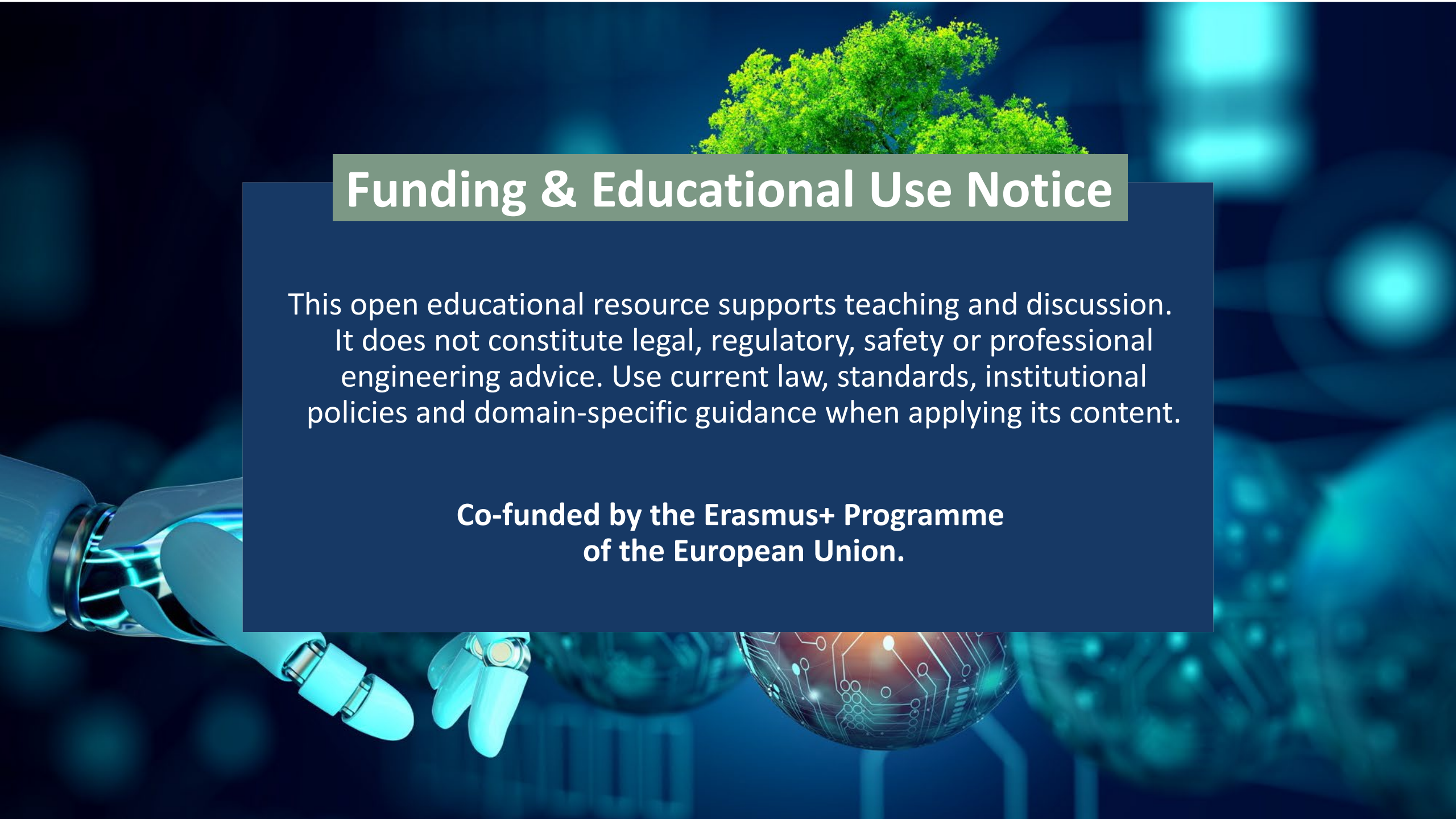
Timing: 5 minutes. Ask learners to point to an artefact that demonstrates each outcome. Record any outcome requiring further practice.

Course Dossier Contribution

Add the following Module 3 evidence to your ethical engineering dossier:

1. Completed data-lifecycle and data-flow map.
2. Purpose, necessity and prohibited-use specification.
3. Privacy architecture and PET selection.
4. DPIA screening record and unresolved specialist questions.
5. Rights, deletion, incident and monitoring controls.
6. Privacy engineering decision record and reflection.

This evidence connects with accountability, values-based design and sustainability in later modules.

A green tree is positioned at the top center of the image. Below it, a dark blue rectangular box contains the title and main text. At the bottom, a robotic hand is visible, holding a glowing globe that features circuitry patterns. The background is a deep blue with some blurred light effects.

Funding & Educational Use Notice

This open educational resource supports teaching and discussion. It does not constitute legal, regulatory, safety or professional engineering advice. Use current law, standards, institutional policies and domain-specific guidance when applying its content.

**Co-funded by the Erasmus+ Programme
of the European Union.**



Ethical Engineer

www.ethicalengineer.eu

Thank You

You have completed **Module 3**



Co-funded by
the European Union