

Module 1

Introduction to AI & Ethics in Engineering Practice



Co-funded by
the European Union



Ethical Engineer

www.ethicalengineer.eu

CONTENTS

01 Ethics & Engineering

02 Justice

03 Autonomy

04 Transparency

05 Non-Maleficence

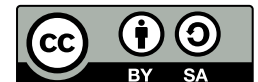
06 Group Activity

07 Quiz



Co-funded by
the European Union

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.



Module Orientation

Purpose

Introduce ethical reasoning as part of AI and robotics engineering practice.

Audience

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

Recommended Duration

Core route 180 minutes. Extended route 225 minutes. Deliver as one class or two shorter sessions.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.

Module Orientation

Learning Sequence

Explain ethics in engineering practice, apply justice, autonomy, transparency and non-maleficence, then analyse an engineering case and justify a decision.

Learner Outputs

Ethical issue map, case analysis, safeguards, engineering ethics decision record and outcome check. These artefacts show how ethical reasoning changes requirements, tests, oversight and approval decisions.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.

Learning Outcome

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

1. Explain how autonomy, justice, transparency and non-maleficence apply to AI and robotics engineering.
2. Identify ethical risks across problem framing, data, model, interface, deployment and monitoring.
3. Analyse an engineering case by connecting design decisions to stakeholder impacts.
4. Propose technical and governance measures that reduce risk and support human oversight.
5. Justify an engineering decision with reference to the EU Ethics Guidelines for Trustworthy AI and relevant DigComp competences.

[Sources]

Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://publications.jrc.ec.europa.eu/repository/handle/JRC128415>

Ethical Reasoning Belongs in Engineering Design

01

Problem
Framing



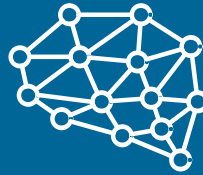
02

Data



03

Model



04

Interface &
Oversight



05

Deployment



Engineering choices shape ethical outcomes throughout the AI system lifecycle:

Ethical reasoning is a design competence. It helps engineers translate European values into technical requirements, testing criteria and accountable decisions

[Sources] - Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70- <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

- **Problem Framing**
Whose need is prioritised?

- **Data**
Who is represented, missing or exposed?

- **Model**
Which errors and trade-offs are accepted?

- **Interface & Oversight**
Who can understand, question or override outputs?

- **Deployment**
Which safeguards, monitoring and incident routes are in place

Alignment with European Competence Frameworks

European frameworks connect ethical principles with practical engineering competence.

Guidance

The EU Ethics Guidelines define trustworthy AI requirements. DigComp 2.2 and DigCompEdu support responsible learner and educator competence.

Application

Learners examine oversight, fairness, privacy, accountability, explainability and safety within an engineering system.

Evidence

Case analysis, safeguards and the decision record show what learners can analyse, test and justify.

01

ETHICS & ENGINEERING



Ethics in Engineering Practice

Ethics is the systematic examination of what is right, fair and responsible.

In engineering, ethics guides professional judgement when technical choices affect people, rights, safety, access or the environment.

An ethical engineer asks:

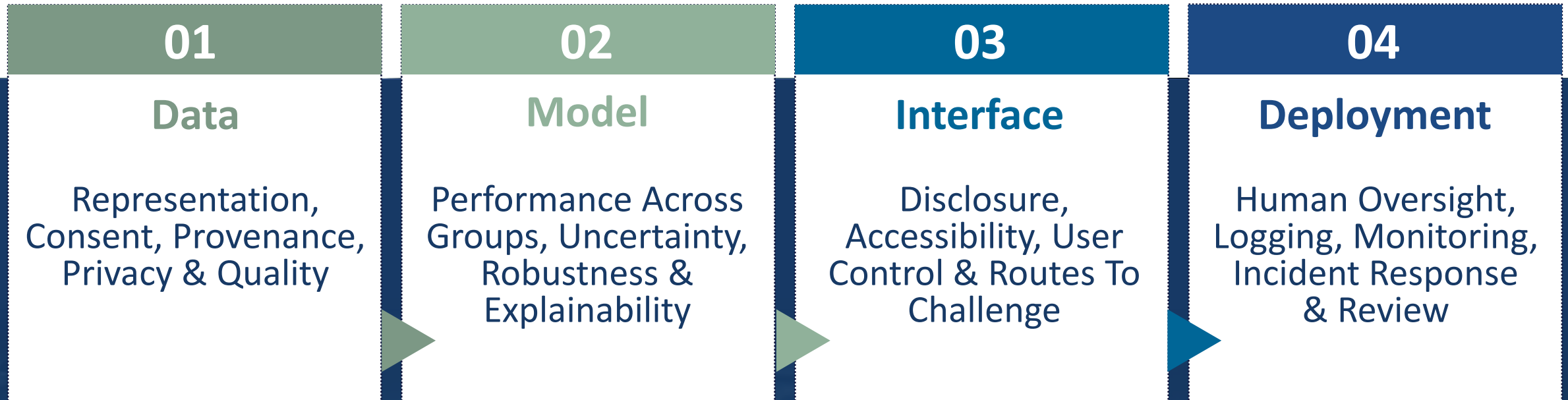
- Who may benefit or be harmed?
- Which values and rights are affected?
- What evidence supports this design choice?
- Who is accountable if the system fails?

Professional codes, institutional policies and law provide requirements. Ethical reasoning supports responsible judgement where several choices remain possible.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

Ethics becomes an engineering requirement

Ethics can be translated into concrete design questions and acceptance criteria:



The engineering task is to make values operational through specifications, tests, documentation and governance.

Ethical Reasoning Uses Several Lenses

The three lenses reveal different parts of the same decision.

- **Consequences**
Examine likely benefits and harms, who experiences them, their probability and whether they can be reversed. A design with strong average performance may still create severe harm for a smaller group.
- **Duties & Rights**
Ask which obligations apply even when another option appears efficient. Human dignity, privacy, non-discrimination, informed choice and safety can place limits on system behaviour.
- **Professional Character**
Ask what competent, honest and responsible engineering practice requires. This includes declaring uncertainty, seeking evidence, recording dissent and escalating a risk when the team lacks authority to resolve it.

[Sources] - <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Law, Standards, Codes & Ethics

- Law sets enforceable obligations and prohibited practices.
- Technical standards define agreed methods, processes and evidence.
- Professional codes state duties such as competence, honesty, public safety and accountability. Organisational policies allocate local responsibilities and approval routes.
- Ethical reasoning examines the choices that remain within these boundaries. It asks which option respects people, distributes risk fairly and can be defended with evidence.
- Compliance leaves many choices about objectives, data, thresholds, interfaces, oversight and monitoring. Engineers remain responsible for recognising and escalating ethical risk within their role.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
Ethical Engineer approved KA220 HED application, WP3 Teaching and Learning, pages 61 to 70.

AI & Ethics (I)

AI can embed or amplify biases, obscure accountability, impact safety and rights, and shift power between stakeholders. While AI offers clear benefits, it also raises ethical challenges that cannot be ignored.

Algorithms can embed or amplify existing biases—for example, a hiring system may unintentionally discriminate against certain groups if trained on biased historical data. Accountability can become obscured when decisions are made by complex models, making it unclear who is responsible if harm occurs.

Safety and rights are also at stake, since errors in autonomous vehicles or medical AI systems can lead to physical damage. At the same time, surveillance technologies may violate privacy and human dignity.

AI & Ethics (II)

Furthermore, AI redistributes power between stakeholders: companies, engineers, regulators, and the public may have competing interests, and those with control over data and algorithms may gain disproportionate influence.

UNESCO's Recommendation on the Ethics of Artificial Intelligence highlights these risks, stressing that human dignity, fairness, transparency, and accountability must remain central in AI development. The EU AI Act echoes these concerns by prohibiting manipulative and exploitative AI practices and imposing strict rules on high-risk systems to safeguard fundamental rights and societal well-being.

AI Ethics Requires a System View

AI systems combine data, models, interfaces, people, organisations and operating environments. Ethical effects can arise from any part of this system. Bias may enter through labels or objectives. Harm may result from a technically accurate output used in the wrong context. A clear explanation may still be ineffective when a person lacks time or authority to act.

AI also changes scale and power. One model can affect many people, reproduce historical patterns and shape access to work, education, health care or public services. Responsibility therefore extends across design, procurement, deployment, monitoring and incident response. The ethical unit of analysis is the whole service and its consequences for affected people.

[Sources] <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Ethical Principles & Professional Engineering Conduct

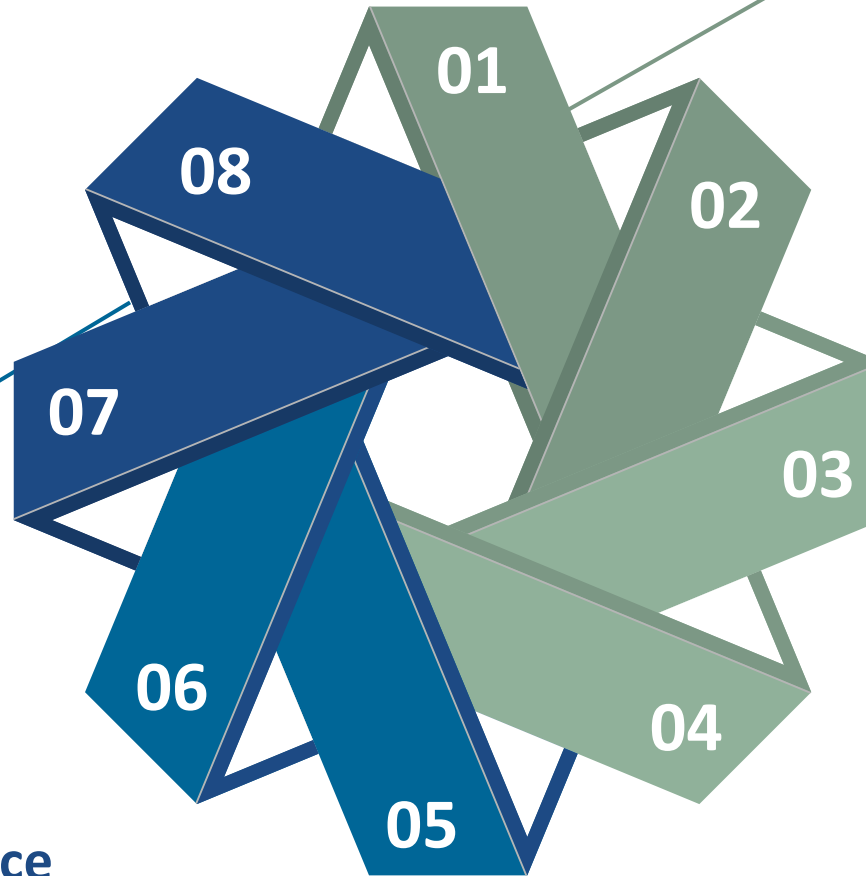
Professional engineering conduct includes:

05. Respect

06. Lawfulness

07. Accountability

08. Commitment to Excellence



01. Honestly
02. Integrity
03. Trustworthiness
04. Fairness

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.
Lapiņa, I. (2019), as cited in the source deck. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Ethical Principles & Professional Engineering Conduct

These values support ethical judgement. AI engineering also needs system-level requirements.

The following sections connect:

- Autonomy
- Justice
- transparency and non-maleficence to data
- model
- Interface
- deployment a
- oversight decisions.



The Four Principles Work Together

Justice asks how benefits, errors and burdens are distributed and whether affected people have a fair process and remedy.

Autonomy asks who has information, choice, authority and meaningful control. Transparency asks which evidence must be available to affected people, operators, engineers and auditors.

Non-maleficence asks which harms are foreseeable and how they will be prevented, detected and addressed. A single design decision can engage every principle. A medical alert threshold influences who receives attention, whether clinicians retain control, what uncertainty is visible and which patients may be harmed by delay or false alarm.

Ethical analysis gains strength when the principles are used as connected questions rather than isolated labels.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Reflect – Discuss - Report (10–15 minutes)

Question

“Where do you already encounter AI (or automation) in daily life or studies? What ethical issues might arise?”

Think **2 mins**

Students write down 2–3 examples/issues.

Pair **5 mins**

Compare lists with a partner.

Share **3-8 mins**

Collect 4–6 examples on the board (for example: face recognition and consent, grading tools and fairness; suggestion systems and transparency).

02

JUSTICE



Justice in AI Engineering

Justice concerns fair & equitable treatment, including how benefits, risks, errors & opportunities are distributed.

For engineers, this requires evidence:

- Are relevant groups represented in the data and validation set?
- Are error rates and impacts assessed across affected groups?
- Do accessibility and exclusion risks shape the design?
- Can people appeal, obtain human review or seek remedy?

Fairness is context-dependent. Teams should document the chosen fairness goals, trade-offs and affected stakeholders.

[Sources] - <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

Justice has Four Engineering Dimensions

1. **Distributive justice concerns** who receives benefits, errors, costs, delays and opportunities. Engineers need outcome measures for relevant groups and operating conditions.
2. **Procedural justice concerns** how decisions are made. People need consistent rules, meaningful human review and a route to challenge an outcome.
3. **Recognition concerns** whether the design understands the needs and experiences of people who may be overlooked, including people with disabilities and minority groups.
4. **Corrective justice concerns remedy.** The service needs a process to correct an error, restore access, investigate causes and prevent recurrence.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://unesdoc.unesco.org/ark:/48223/pf0000381137>

Real Case: Face Recognition Errors

Official evidence:
NIST face
recognition study

NIST evaluated nearly 200 face recognition algorithms using more than 18 million images of over 8 million people. Demographic differentials appeared in the majority of algorithms tested. The error pattern depended on the algorithm, application and image data.

The engineering issue concerns use context as well as model performance. A false match can expose a person to investigation. A failed verification can deny access to a service. Engineers must identify the relevant error, test affected groups and realistic image conditions, document uncertainty and provide human review and remedy.

Discussion:

Which error creates the greatest harm in campus access, policing and phone authentication?

[Sources] <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>

Applied Task: Justice in Infrastructure Allocation

A regional authority uses an AI model to prioritise bridge inspections. Urban road bridges have rich sensor and maintenance data. Rural bridges, pedestrian crossings and routes used by disabled people have sparse records. The model directs most inspection capacity to sites with the strongest data and highest traffic volume.

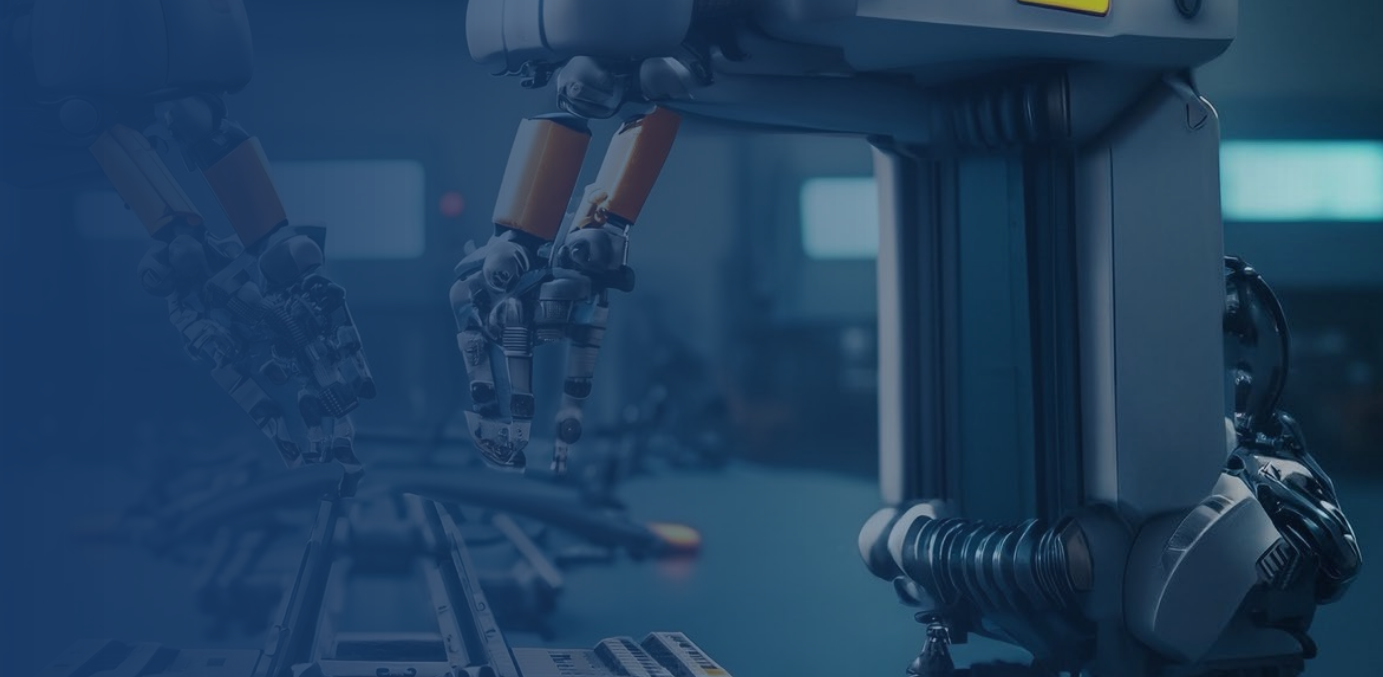
Justice asks how safety attention, risk and access are distributed, and whether overlooked communities have a fair route to review.

Decide: GO, REVISE or STOP. Specify affected groups, missing evidence, a fair allocation rule, manual nomination route, acceptance criteria, accountable owner and remedy.

Quality check: Who receives earlier protection, who carries residual risk, and how can an omission be corrected?

03

AUTONOMY



Human Autonomy & Engineering Responsibility

Human autonomy is the capacity to make informed choices and act without undue pressure. In AI systems, it depends on meaningful information, consent where appropriate, genuine user control and routes to question or override outcomes.

Engineers influence autonomy through interface design, default settings, automation level, override mechanisms and the allocation of decision authority.

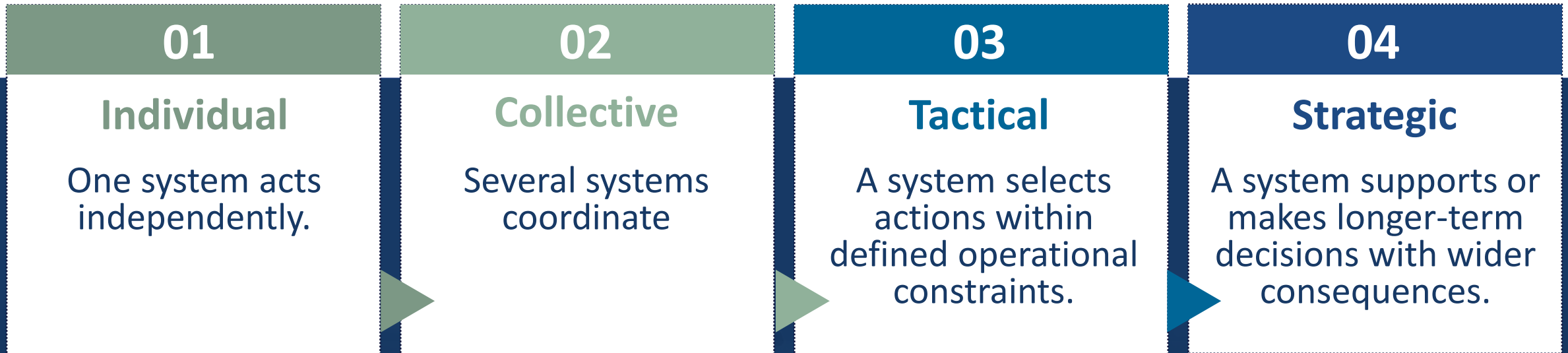
Key Distinction:

System autonomy describes what a machine can do independently. Human autonomy concerns the agency and rights of people affected by the system

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

Levels of System Autonomy

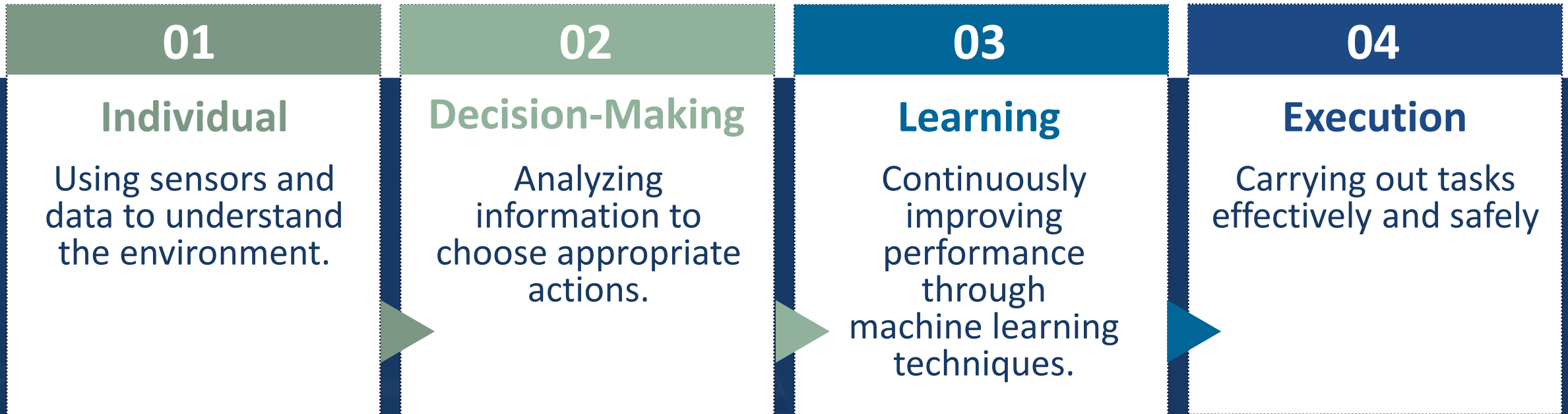
Engineers can describe autonomy at different levels:



For each level, specify operating boundaries, human oversight, fail-safe behaviour, accountability and evidence required before deployment. Each level has a distinct risk profile and safeguard requirement.

Autonomous AI (II)

Key features of autonomous AI include:



Autonomous AI is transforming various industries such as transportation, manufacturing, healthcare, and logistics, offering increased efficiency, safety, and new capabilities. However, it also raises ethical, safety, and regulatory considerations, particularly regarding decision accountability and impact on employment.

Autonomy Changes Decision Authority

For every automated function, engineers should specify the permitted action, operating boundary, human role, information required for intervention, response time, fallback and accountable owner. A person named in a workflow has meaningful control only when the person can recognise a problem, understand the system state and act before harm occurs.

System autonomy describes functions performed without immediate human action. Human autonomy concerns the ability of people to understand, choose, refuse, intervene and challenge decisions. The ethical question is therefore about authority: who may decide, who may act and who can stop the process.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Real Case: Automated Driving Oversight

In the 2018 Tempe crash, an Uber test vehicle struck and killed a pedestrian. The NTSB found inadequate safety risk assessment, ineffective oversight and a weak safety culture. Collision warning and automatic emergency braking had been deactivated. Official investigation: NTSB adopted report.

Human oversight is a designed safety function. It requires clear system-state information, realistic vigilance demands, enough response time, tested intervention, safe fallback and organisational monitoring.

Discussion:

Which system, interface and governance decisions shaped the outcome?

Applied Task: Autonomy in Process Control

An AI controller adjusts cooling-water valves in a hydrogen electrolyser. It can change a valve position every second. During a sensor fault, unsafe pressure can develop within eight seconds.

An operator needs about ten seconds to interpret the alarm and a further five seconds to act. The supplier describes the system as human supervised because an emergency-stop button is available..

Decide: GO, REVISE or STOP. Specify the controller's permitted actions, operating boundary, human authority, alarm information, response time, automatic safe state, independent protection layer, verification test, evidence owner and review trigger.

Quality Check: Can human control operate in time under realistic failure and workload conditions?

04

TRANSPARENCY



Transparency as an Engineering Property

Transparency enables affected people, operators, auditors and regulators to understand what an AI system is for, how it is used and where its limits lie

Engineers create transparency through documentation, traceability and interface design

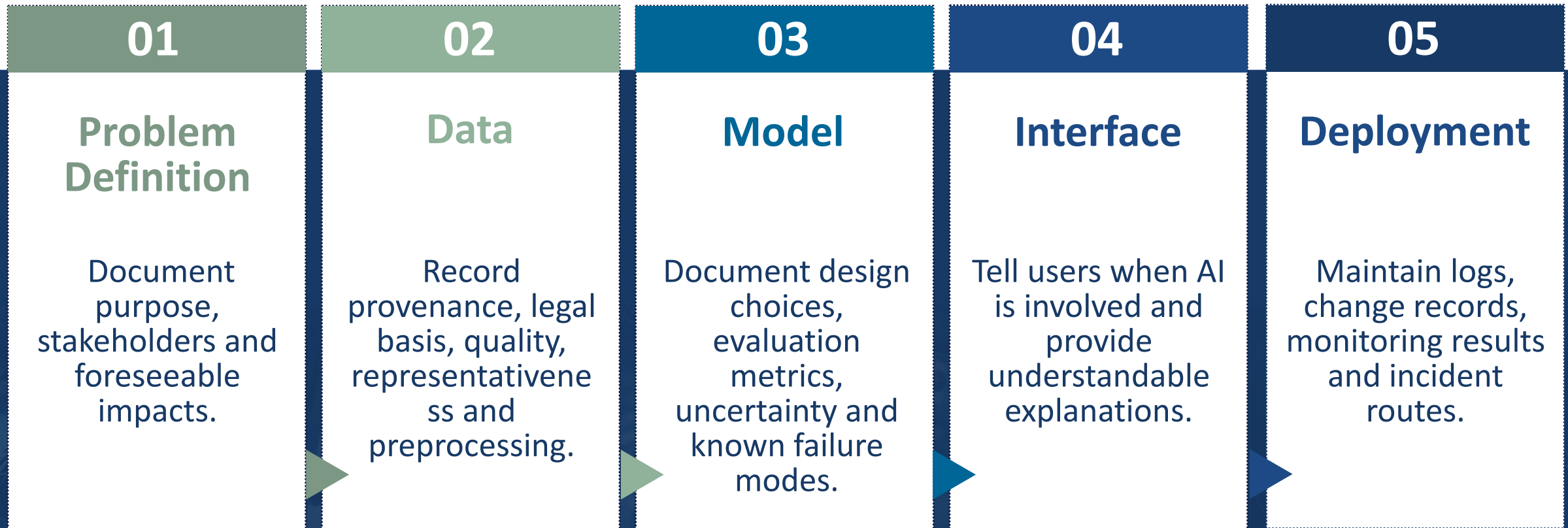
Relevant information can include:

- Intended use and prohibited use
- Data sources, quality and known gaps
- Model limits, uncertainty and performance across groups
- Human oversight and escalation routes
- Monitoring, incidents and system changes

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

Transparency across the AI lifecycle

Transparency should be appropriate to the audience while protecting privacy, security and legitimate intellectual property.



Engineering Trade-offs in Transparency

Transparency supports trust, accountability, testing and informed use. It can create risks if sensitive personal data, security details or protected intellectual property are disclosed without safeguards.

The engineering objective is purposeful transparency:

- Disclose what each stakeholder needs to act safely and challenge decisions
- Protect information that could create privacy or security harm
- Document why information is withheld and who can access it for audit
- Test whether explanations are accurate, usable and accessible

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

Key Components of AI Transparency (I)

- **Explainability**

Ability to explain how the AI arrived at a particular decision or outcome (e.g., why a loan was denied).

- **Data Transparency**

Clear documentation of what data was used to train the model, including sources, quality, and potential biases

- **Algorithmic Transparency**

Access to the algorithms used (e.g., open-source models, documentation of decision logic).

Transparency is Engineering Evidence for action

Transparency supports safe action, challenge and accountability. Its purpose changes with the audience. Operators need system state, uncertainty and the required response. Affected people need to know that AI was used, how an outcome affects them and how to seek review. Engineers and auditors need traceability across data, tests, versions, decisions and incidents.

Engineering artefacts make this practical:

- System and model records define intended use, limits and failure modes.
- Test evidence covers relevant conditions and affected groups.
- Interface messages communicate uncertainty, action and escalation.
- Logs, change records and incident reports support reconstruction and review.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

Transparency depends on audience and purpose

Useful transparency enables a specific person to understand, act, challenge or audit. Publishing technical detail without identifying the audience can create apparent openness while leaving the decision unaccountable. The information should be accurate, timely, accessible and linked to a real route for action.

- An affected person needs to know that AI is involved, the decision it supports, relevant reasons, important limits and how to obtain human review.
- An operator needs system state, confidence, required action, operating limits and escalation instructions.
- Engineers and auditors need data provenance, versions, tests, logs, incidents, changes and approvals.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Real Engineering Case: Boeing 737 MAX Transparency

Official investigation: [Boeing 737 MAX report materials.](https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation)

[Sources] <https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation>

The United States House Committee investigation into the Boeing 737 MAX examined the design, development and certification of the aircraft. Its report identified repeated failures involving the Maneuvering Characteristics Augmentation System, known as MCAS, and wider weaknesses in certification and oversight.

The transparency issue concerns information needed for safe action and independent review. Pilots, airlines and regulators required accurate evidence about system authority, sensor dependence, failure behaviour, cockpit indications, procedures and training needs. Incomplete or fragmented information weakened their ability to understand and challenge the design.

Discussion:

Which engineering records, hazard evidence and change decisions should have been visible, to whom, and at what approval point?

Applied Task:

Transparent Rail Engineering Procurement

A rail operator plans to procure an AI system that identifies track defects from inspection images. The vendor reports 97 per cent overall accuracy but will not disclose performance by rail type, weather, camera condition or defect severity. Training-data provenance and known failure modes are unavailable. The contract permits model updates without notice. Maintainers remain accountable for missed defects.

Write

Write the evidence and contract requirements needed before procurement. Cover intended use, operating limits, dataset provenance, defect-level results, uncertainty, version and change records, access to logs, incident cooperation, independent testing and exit rights.

Decision

State one condition that requires REVISE and one condition that requires STOP. Explain how each audience can act on the information supplied

05

NON- MALEFICENCE

Non- Maleficence:

Not to harm, anticipate, and mitigate risks. The principle of non-maleficence requires engineers to avoid creating systems that cause physical, psychological, or societal harm.

It also calls for anticipating potential dangers in advance and putting safeguards in place to mitigate unintended consequences.

Harm Extends Beyond Physical Injury

- Physical harm includes injury and unsafe operation. Psychological harm includes distress, manipulation and loss of confidence. Economic harm includes lost work, higher costs and exclusion from services. Rights-based harm includes privacy intrusion, discrimination and denial of due process. Social and environmental harm includes surveillance, unequal power, resource use and effects on communities.
- Non-maleficence requires engineers to identify exposed people and plausible harm before selecting controls. Severity, likelihood, exposure and reversibility should be considered together. A rare severe harm, a frequent low-level burden and an error that cannot be corrected each require a different response.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://unesdoc.unesco.org/ark:/48223/pf0000381137>

Safety is a Lifecycle Obligation

- Before release, define intended and prohibited use, identify hazards and foreseeable misuse, test edge cases and degraded conditions, and specify measurable acceptance criteria. During operation, monitor performance, group outcomes, overrides, complaints, incidents and context change. After an incident, protect affected people, preserve evidence, determine causes, correct controls and verify the change.
- Controls should address the source of harm where possible. Teams can change the objective, remove a hazardous function, restrict the operating domain, add technical detection and fallback, strengthen human review or stop deployment. A safety case remains valid only while its assumptions, system version, context and evidence remain valid.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Real Engineering Case: BP Texas City Process Safety

On 23 March 2005, explosions at the BP Texas City refinery killed 15 people and injured 180. The United States Chemical Safety Board reported that a distillation tower flooded, became overpressurised and released hydrocarbons from an atmospheric vent stack. Its investigation identified organisational and safety deficiencies across the corporation.

Non-maleficence requires engineers to prevent foreseeable harm through the whole system. Equipment design, alarms, operating procedures, staffing, management of change, site layout, leading safety indicators and executive oversight all contribute to risk. Personal safety statistics cannot substitute for evidence about catastrophic process hazards.

Discussion:

Which prevention layers failed, what evidence should have triggered action, and who had authority to stop the restart?

Applied Task: Bounded AI Process-Safety Pilot

A chemical plant proposes an AI advisory tool that predicts compressor surge and recommends bypass-valve changes. Validation covers steady-state operation. There is no evidence for start-up, shutdown, sensor drift or loss of communications. The recommendation is applied automatically unless the operator cancels it within six seconds. The supplier proposes a three-month live pilot

Decide: GO, REVISE or STOP. Define the pilot boundary, excluded operating modes, independent protection layers, human response time, pre-release tests, monitoring signals, incident route, shutdown thresholds, evidence owner and review date.

Produce a concise safety case: Hazard, causal path, control, verification evidence, residual risk and accountable approval. State which missing evidence prevents live operation.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
<https://www.csb.gov/bp-america-texas-city-refinery-explosion/>

06

GROUP ACTIVITY



Case A: BridgeSense Inspection Prioritisation

BridgeSense ranks bridges for inspection using images, vibration sensors, traffic volume and maintenance history. Older rural bridges have fewer sensors and incomplete records. Only the top-ranked 20 sites will be inspected. A low ranking may reflect missing data rather than safety. Delay could cause harm, road closure or loss of essential access.

Task:

Decide GO, REVISE or STOP. State how the ranking should inform professional judgement and resource allocation.

Evidence Available:

Overall accuracy and urban validation.

Evidence Missing:

Results by bridge type and condition, uncertainty, community impact, override and review records

Case B: WindControl Turbine Operation

WindControl adjusts blade pitch and maintenance timing. It was trained on newer turbines in moderate weather. The site also has older turbines, icing, strong gusts and intermittent communications. The interface does not show confidence or supporting data.

A false alarm may cause shutdown. A missed fault may damage equipment or endanger technicians and communities

Task:

Approve a limited pilot, require redesign or stop. Define operating limits, human authority and conditions for automatic action.

Evidence Available:

Extreme-weather tests, older-turbine results, response time, fail-safe behaviour, change control and near-miss monitoring.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Engineering Case Analysis & Team Output (30 minutes)

Choose one case.

In groups of four or five, assign systems, safety, evidence, stakeholder and recorder roles

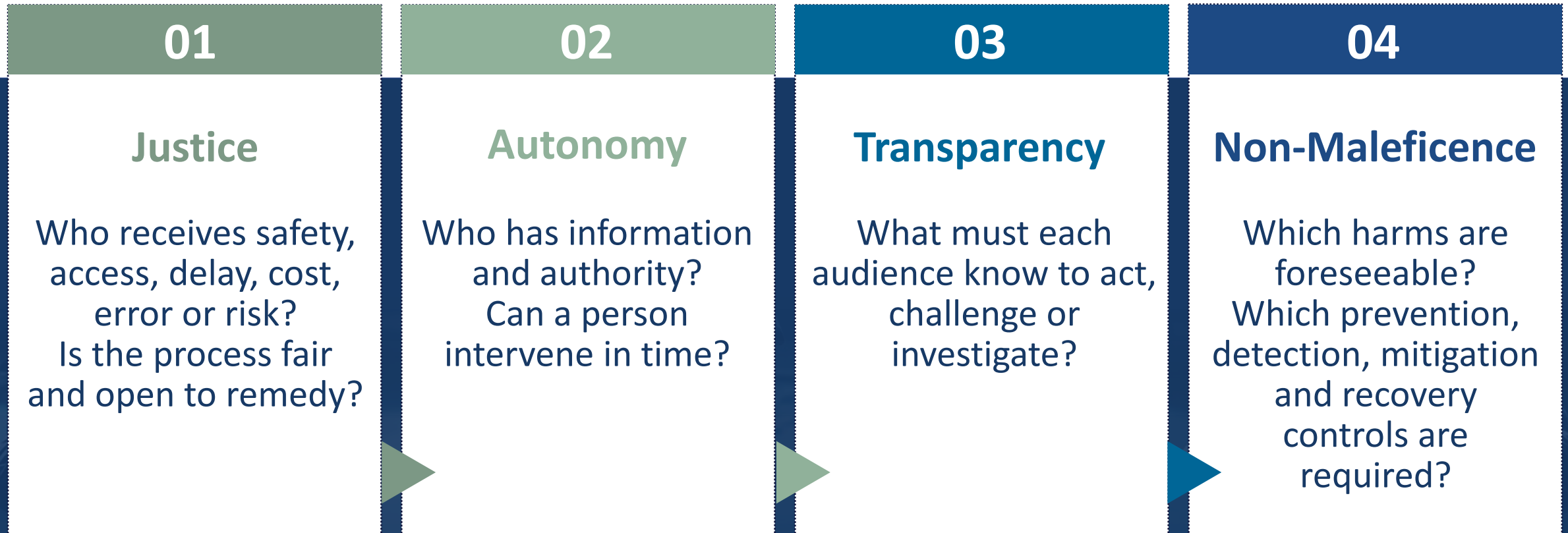
1. Define intended use, system boundary, conditions and affected people.
2. Name the engineering decision that creates the dilemma.
3. Trace how data, model, interface, automation and monitoring influence outcomes.
4. Apply all four principles and identify any conflict.
5. Propose technical and governance safeguards, with tests and owners.
6. Record GO, REVISE or STOP and the review trigger.

Output: one decision record and a two-minute evidence-based justification.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

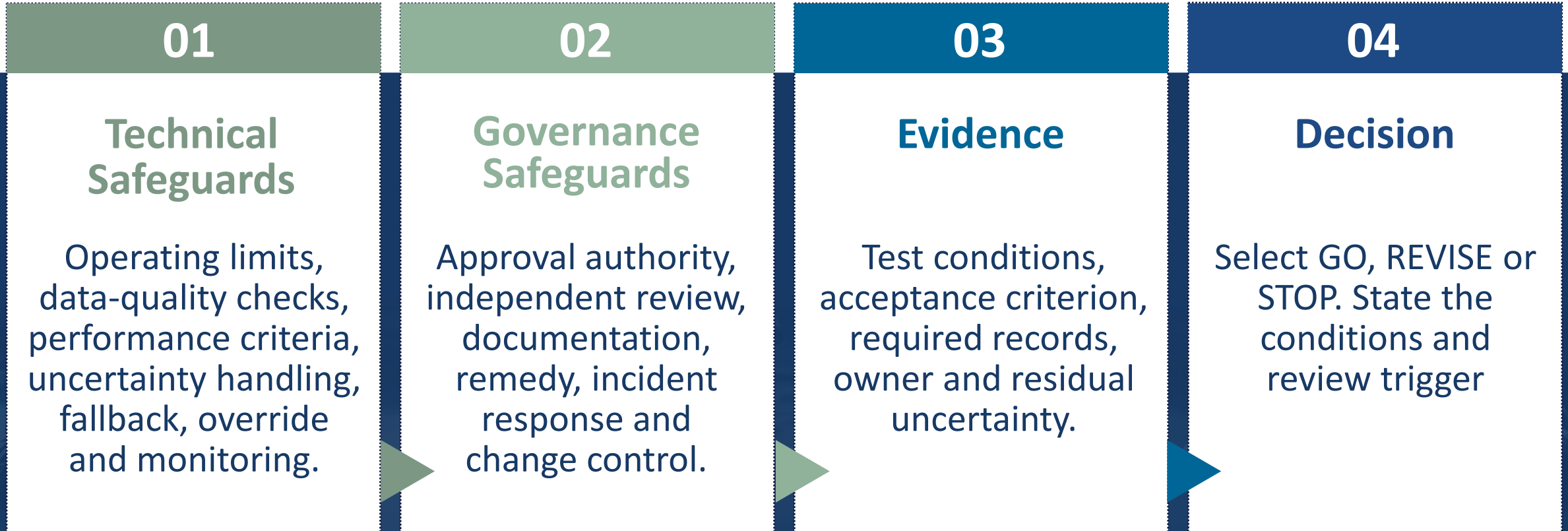
Apply the Four Principles to one Engineering Decision

Use the same critical decision throughout so that every principle remains connected to engineering action.



Design Safeguards, Evidence & Oversight

Turn the ethical analysis into an engineering proposal.



Worksheet: Engineering Ethics Decision Record

Purpose: Integrate the four principles into one assessable engineering decision. Complete the record for BridgeSense or WindControl, then prepare a one-minute justification. A strong record links principles to decisions, safeguards, evidence, an accountable owner and a review trigger.

System, intended use and affected people:

Critical engineering decision:

Justice, autonomy, transparency and non-maleficence:

Technical safeguard and acceptance test

Governance safeguard and accountable owner

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Continues On Next Slide



Missing evidence, uncertainty and residual risk:

Decision: GO REVERSE STOP

Conditions, monitoring measures and review trigger:

Dissent or alternative view:

Justification: Why does the evidence support this decision?

Role-Play Scenario

(25 Minutes)

Scenario:

The university is considering autonomous patrol robots for night-time monitoring. Cameras and person detection promise quicker incident response. Students raise concerns about privacy, discrimination and false alerts.

Roles, one per group:

- **Engineering lead:** safety, performance and operating limits
- **Data scientist:** datasets, accuracy and bias
- **Student representative:** rights, inclusion and personal data
- **Facilities manager:** operations and liability
- **Ethics or legal adviser:** risk assessment and documentation

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

Individual Reflection & Transfer

(15 Minutes)

Reflection completes the group activity and transfers the reasoning to another engineering context.

Write one or two paragraphs:

- Which two principles most influenced your decision, and why?
- Which trade-off or uncertainty was hardest to resolve?
- Which engineering change is required before deployment?
- What evidence would cause you to revise the decision?
- How would the reasoning change in another discipline?

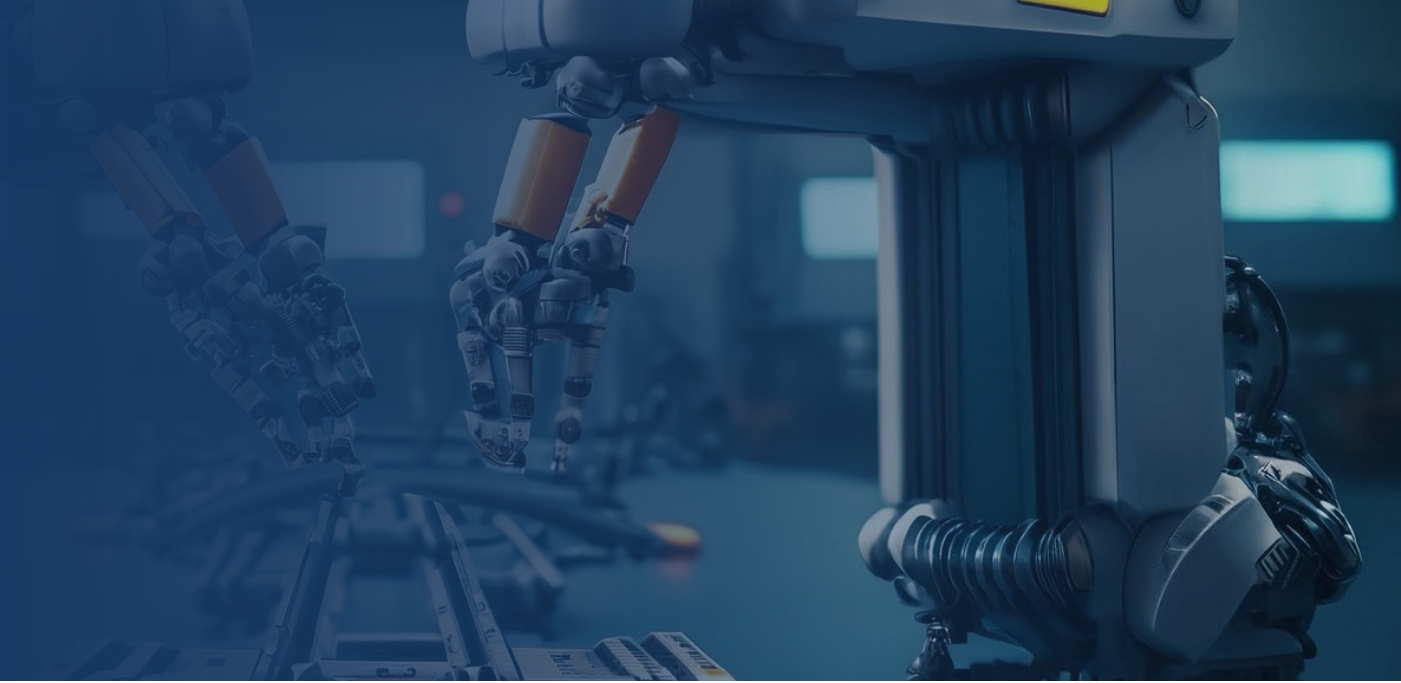
Evidence:

A concise argument connecting principles, an engineering decision, safeguards and evidence. Written, audio or recorded responses are acceptable.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.

07

QUIZ



Quiz

1. Define *transparency* in AI and give one engineering practice that supports it.
2. Which principle aims to **avoid harm**, and name one risk-mitigation method in AI projects.
3. List **two ways** biased training data can affect model outcomes.
4. In an AI-assisted hiring tool, identify **one decision point** where engineers can reduce unfairness.
5. A robot misclassifies mobility aids as obstacles, thereby blocking access. Which principle is most directly violated?
 - a. *Autonomy*
 - b. *Justice*
 - c. *Transparency*
 - d. *Non-maleficence*

Quiz Answers & Teaching Rationale

1. Transparency means relevant clarity, traceability and communication. Supporting practices include dataset documentation, model records, explanations, logs and change records.
2. Non-maleficence. A valid method could be hazard analysis, representative testing, operating restrictions, safe fallback, monitoring or incident response.
3. Biased data can create unequal error rates, poor performance for under-represented groups, harmful proxies and reinforcement of historical patterns.
4. Valid decision points include data selection, labels, features, objective, threshold, interface, human review and monitoring.
5. B, Justice. Blocking access distributes a burden unfairly. Non-maleficence may also apply because the error causes harm. Credit a reasoned answer.

[Sources] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Learning Outcome Check

Before finishing, students should be able to:

- Name the engineering decision that created or reduced harm in a case.
- Use at least two ethical principles to explain the issue.
- Recommend one technical safeguard and one governance safeguard.
- State what evidence, documentation and monitoring would be required.
- Explain how affected people can understand, question or challenge the outcome.

Evidence of learning:

Group analysis, role-play decision, reflection response and quiz.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.

Core References & Further Reading

- European Commission High-Level Expert Group on AI. [Ethics Guidelines for Trustworthy AI](#).
- Vuorikari, Kluzer and Punie. [DigComp 2.2](#).
- **UNESCO.** [Recommendation on the Ethics of Artificial Intelligence](#).
- European Union. [Artificial Intelligence Act](#).
- **NIST.** [Face Recognition Vendor Test: demographic effects](#).
- NTSB. [Tempe automated vehicle crash investigation](#).
- United States House Committee. [Boeing 737 MAX investigation](#).
- United States Chemical Safety Board. [BP Texas City refinery investigation](#).
- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning, pp. 61-70.

[Sources]

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

<https://unesdoc.unesco.org/ark:/48223/pf0000381137>

<https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition>

<https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation>

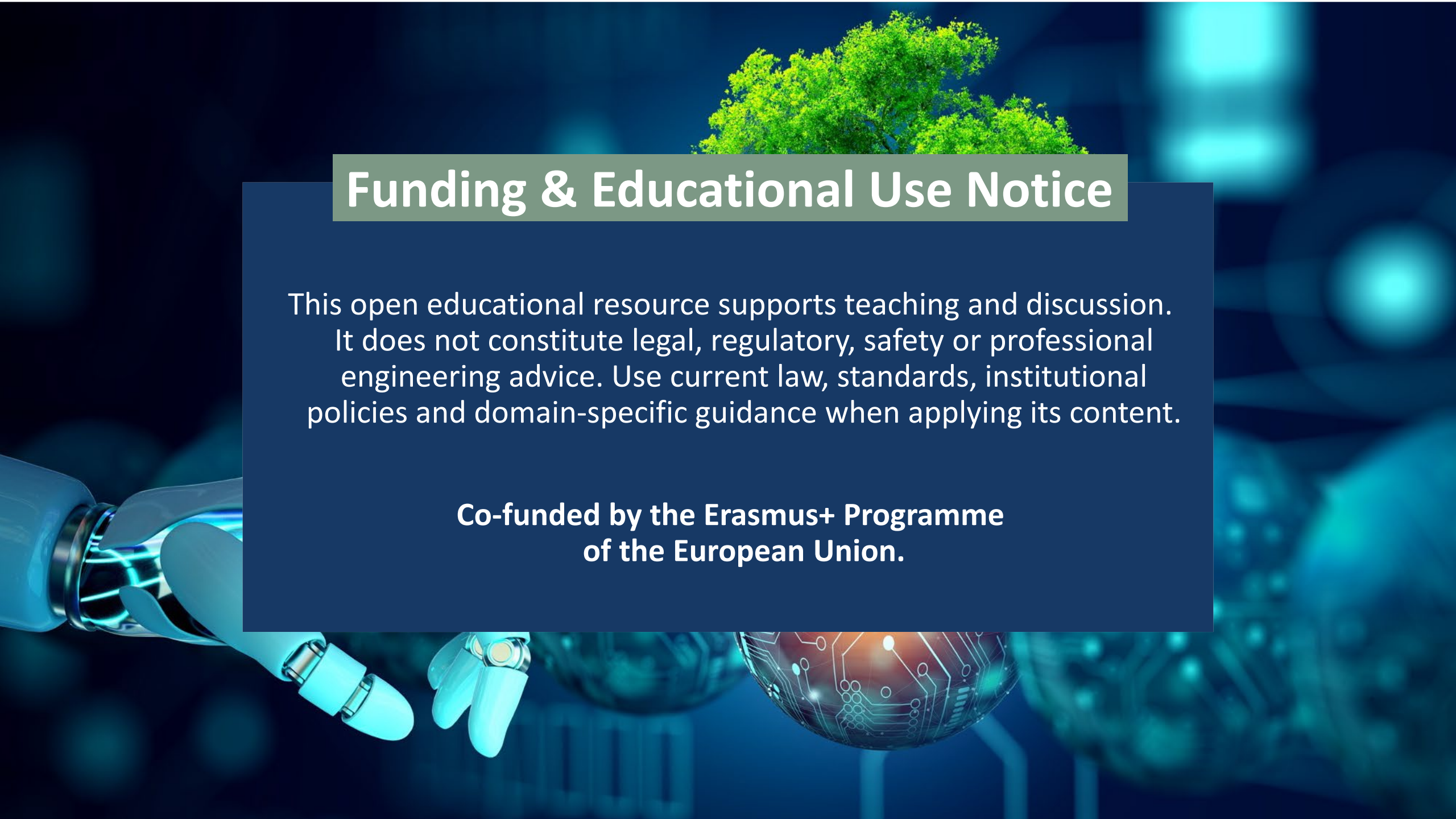
Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning (OER Development), pp. 61-70.

<https://publications.jrc.ec.europa.eu/repository/handle/JRC128415>

<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

[softwarehttps://www.nts.gov/investigations/AccidentReports/Reports/HAR1903.pdf](https://www.nts.gov/investigations/AccidentReports/Reports/HAR1903.pdf)

<https://www.csb.gov/bp-america-texas-city-refinery-explosion/>

A green tree with dense foliage is positioned at the top center of the image. The background is a deep blue with a subtle, glowing circuit or network pattern. In the lower-left corner, a robotic arm with blue and silver segments is visible, reaching towards the center. The overall aesthetic is high-tech and futuristic.

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 1**



Co-funded by
the European Union