Module 4

Engineering Responsibility & Accountability in Autonomous Systems

Ethical Engineer

Co-funded by the European Union

www.ethicalengineer.eu

CONTENTS

01 Allocate Responsibility & Authority

02 Design Resilience & Human Control

03 Learn Accountability Through Cases

04 Produce a Traceable Engineering Decision

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.

CC BY SA

Ethical Engineer

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	Energy, water and hardware Societal and lifecycle impacts Metrics, governance and redesign

Module Orientation

AI failures can cause real-world harm. This module shows how engineers design transparency, safety and accountability before operation.

Purpose

Equip engineering students to allocate responsibility, analyse failure and defend safety related decisions with traceable evidence.

Audience

Higher education students in engineering, computing, robotics and safety related disciplines.

Builds On

Modules 1 to 3: ethics, transparency, fairness, privacy, data governance and engineering decision records.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.

Module Orientation

AI failures can cause real-world harm. This module shows how engineers design transparency, safety and accountability before operation.

Duration

Core route 210 minutes. Extended route 270 minutes.

Learning Sequence

Allocate authority, design resilience, analyse cases and produce an accountable engineering decision.

Learner Outputs

Responsibility map, failure analysis, assurance argument and accountability record.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.

Learning Outcome

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

1. Distinguish responsibility, accountability and legal liability.
2. Map decision and control ownership across an autonomous system lifecycle.
3. Analyse data, model, operational, organisational and human factor failure modes.
4. Specify resilience, fallback, handover, logging, assurance and change controls.
5. Produce an accountability record linking evidence, owners, residual risk and review routes.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.

Accountability must be Designed Before Failure

- Autonomous systems distribute decisions across data suppliers, model teams, integrators, operators, managers and regulators. When authority is unclear, a safety issue can remain unaddressed.
- Accountability requires a person or organisation to explain and justify a decision, provide evidence, accept review and support correction or remedy.
- Engineers make this possible through decision rights, traceability, test evidence, monitoring, incident routes and authority to stop operation.

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

European Frameworks Connect Duties with Engineering Evidence

Trustworthy AI guidance identifies accountability, human agency and technical robustness as connected requirements.

The EU AI Act assigns duties to defined actors and requires relevant high risk systems to support risk management, logging, human oversight, accuracy, robustness and cybersecurity. The Product Liability Directive addresses liability for defective products, including software, under its conditions.

Frameworks do not allocate every project task. Engineering governance must define who decides, implements, verifies, approves, monitors and responds.

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

01

Allocate Responsibility & Authority

Distinguish duties, map decision rights and connect ethical requirements with evidence.

Responsibility, Accountability & Liability are Distinct

- Responsibility concerns an assigned duty or action and can be shared across roles. One engineer may run a test; another may verify the result.
- Accountability is the obligation to explain and justify a decision, provide evidence and accept review. A named release owner must be able to defend the outcome.
- Legal liability concerns consequences determined under applicable law and facts. In a hospital, many staff contribute to care, but authority, decision records and review routes must show who could approve, stop or correct the action.

[Sources]
- https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727
- https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng

Map Accountability Across the System Lifecycle

Choose one safety related decision. At each stage, identify authority, responsibility, evidence, control and escalation.

01	02	03	04	05	06
Purpose & Boundary	**Requirements & Design**	**Integration & Verification**	**Release & Approval**	**Operation & Oversight**	**Incident & Change**
What function, operating domain and unacceptable outcome define the decision?	Who owns requirements, hazard controls, interfaces and supply chain assumptions?	Who integrates components, verifies controls and resolves failed evidence?	Who can accept residual risk, approve release or require revision?	Who monitors performance, acts on alerts and has stop authority?	Who investigates, preserves evidence, approves restart and controls updates?

OUTPUT An accountability map linking each stage to a decision owner, control owner, evidence owner, stop authority and review trigger.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai - https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727

Current EU Liability Context Requires Precise Claims

- The EU AI Act regulates defined risks and actor obligations. It does not create a complete civil liability code for every AI related harm.
- Directive (EU) 2024/2853 updates product liability rules and includes software within the product framework. National civil law, contracts, data protection, consumer protection, safety rules and sector legislation may also apply.
- The proposed AI Liability Directive was withdrawn in 2025 and did not become EU law. Preserve evidence and obtain qualified legal advice for real cases.

[Sources]

- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727
- https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng
- https://eur-lex.europa.eu/procedure/EN/2022_303

Decision Rights Turn a Role List into Governance

- Define who can set an operating limit, accept a risk, approve a model update, override automation, suspend use and communicate an incident.
- Separate preparation, verification and approval where independence matters. Record consultation and dissent. Avoid assigning accountability to a committee without a named decision owner.
- Test the map with a realistic failure and ask who acts in the first minute, hour and day.

[Sources]
- https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727

Providers, Deployers, Operators & Suppliers See Different Evidence

- A provider controls design and development evidence. A deployer controls the use context, configuration, staff and operational process. An operator acts through the interface. A supplier may control model, software or hardware evidence.
- Technical documentation and agreements should connect these views through intended use, limits, tests, logs, updates, incidents and review access.
- A missing evidence route can make accountability impossible even when roles are named.

[Sources]

- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727
- https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng

Risk Ownership Requires Authority, Resources & Escalation

- A risk owner monitors the risk and decides how it is managed. A control owner implements a safeguard. An evidence owner maintains the records needed for verification and review.
- Assigning ownership without budget, competence or authority creates a paper control.
- For each significant risk, record the owner, control, acceptance criterion, residual risk, escalation threshold and review date.

[Sources]

- https://www.nist.gov/itl/ai-risk-management-framework
- https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Ethical Principles Become Requirements & Usable Evidence

- Autonomy becomes human authority, information and response time. Justice becomes distributional testing, fair process and remedy. Transparency becomes audience specific evidence and traceability. Non maleficence becomes hazard controls, verification and monitoring.
- **Collect evidence that supports the decision:** requirements, hazard analysis, test results, model cards, datasheets, approval records, logs, changes and incidents.
- A reviewer should be able to trace an affected outcome to the system version, data boundary, condition, control, human action and approval.

[Sources]

- https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727

Worked Mechanical Engineering Example: Pump Release Decision

A maintenance team proposes returning a centrifugal process pump to full load after bearing replacement. An AI condition monitor supports the decision, yet release authority remains an engineering responsibility.

01

Critical Decision

Release the pump for continuous operation at 3,600 rpm. The maintenance manager owns the release decision.

02

Responsible Work

The mechanical engineer sets vibration and temperature limits. The technician records alignment, torque and bearing installation.

03

Independent Evidence

A separate reliability engineer reviews baseline vibration, thermal trend, alignment report and trip test before approval.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://www.nist.gov/itl/ai-risk-management-framework

Worked Mechanical Engineering Example: Pump Release Decision

A maintenance team proposes returning a centrifugal process pump to full load after bearing replacement. An AI condition monitor supports the decision, yet release authority remains an engineering responsibility.

04 Controls

Use a controlled speed ramp, automatic high vibration trip, guarded exclusion zone and local manual shutdown.

05 Stop Authority

The control room operator and safety lead can stop the pump. The supplier cannot override site shutdown authority.

06 Decision & Review

REVISE until the trip test and independent review pass. Approve a 24 hour monitored run, then review trends before full release.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://www.nist.gov/itl/ai-risk-management-framework

Worked Mechanical Engineering Example: Pump Release Decision

A maintenance team proposes returning a centrifugal process pump to full load after bearing replacement. An AI condition monitor supports the decision, yet release authority remains an engineering responsibility.

Accountability Changed the Release

A model score became one item in a versioned evidence chain with independent verification, named stop authority and staged approval.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://www.nist.gov/itl/ai-risk-management-framework

02

Design Resilience & Human Control

Specify fault tolerance, fallback, handover, bounded authority, logging and change control.

Fault Tolerance Starts with a Defined Fault.

Fault tolerance allows a system to continue within safe limits after a defined fault. Start by naming relevant sensor, actuator, data, software, communication and human action faults.

Engineering evidence must show which critical function is protected, how the fault is detected and which transition keeps the system safe.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://www.nist.gov/itl/ai-risk-management-framework

Fault Tolerance Starts with a Defined Fault

01

Fault Model

Define credible single, combined and common cause faults, their duration and the safety function they can affect.

02

Control Architecture

Use independent redundancy, diversity, monitoring, isolation and a safe state suited to the hazard.

03

Verification

Test detection coverage, transition time, degraded operation, operator response and residual single points of failure.

A design is fault tolerant only for the faults and operating conditions supported by evidence. Unknown dependencies and untested combined faults remain residual risk.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning. https://www.nist.gov/itl/ai-risk-management-framework

Redundancy Needs Independence & a Decision Rule

- Two components offer weak protection when they share power, data, software defects or environmental vulnerability. Define how disagreement is detected and resolved.
- Use diverse sensors or methods where common cause failure matters. Verify coverage, latency and transition after a disagreement.
- Document the protected function, assumed independence, evidence and remaining single points of failure.

[Sources]

- https://www.nist.gov/itl/ai-risk-management-framework

Detection, Containment & Recovery Form One Control Chain

- Detection identifies abnormal behaviour. Diagnosis distinguishes likely causes. Containment prevents spread. Fallback moves to a safer function or state. Recovery restores service under controlled conditions.
- Each stage needs a time requirement, observable signal and owner. A warning without a feasible response is weak protection. A fallback that relies on the failed component is not independent.
- Test the complete chain under realistic load and degraded communication.

[Sources]

- - https://www.nist.gov/itl/ai-risk-management-framework

Graceful Degradation Preserves the Most Important Functions

- Reduce capability in a controlled way when confidence, sensing, communication or computation degrades. Define which functions remain, which stop and how users understand the change.
- Examples include lower speed, reduced automation, conservative thresholds, local control or safe shutdown.
- Verify that the degraded mode is stable, observable and suitable for the environment. Record its trigger and maximum permitted duration.

[Sources] - https://www.nist.gov/itl/ai-risk-management-framework

Human Oversight is a Timed System Function

- Meaningful oversight requires a person to recognise the situation, understand system state, decide and act before harm. Response time includes perception, interpretation and physical action under realistic workload.
- Design alerts, uncertainty display, authority, override, fallback and training as one function. Continuous monitoring of rare events is unreliable without support.
- Test with representative operators, conditions and failure scenarios.

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

Handover Must Specify State, Time & Authority

- A handover changes control between automation and a person or another system. The receiving actor needs current state, reason, available actions, remaining time and confirmation that control has transferred.
- Pre alerts can prepare the operator. A minimum risk manoeuvre may be required when no response occurs. Interfaces should prevent ambiguous shared authority.
- Log the trigger, information presented, response and resulting state.

[Sources]

- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727
- https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2

Agentic Systems Need a Bounded Action Envelope

- Systems that plan, call tools or act across connected services can create long action chains. Define permitted goals, data, tools, physical states, spending or resource limits and prohibited actions before deployment.
- Require approval gates for safety critical or irreversible actions. Monitor tool calls, state changes and failed constraints. Provide rapid revocation, a safe fallback and a clear operator view.
- Accountability evidence must show who authorised the envelope, which version acted and how the action was stopped or reversed.

[Sources]

- https://www.nist.gov/itl/ai-risk-management-framework
- https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727

Logs Make Reconstruction & Learning Possible

- Useful logs connect time, system version, configuration, inputs, outputs, confidence, alarms, human actions, overrides and changes. They should support incident response without collecting unnecessary personal data.
- Define structured format, clock synchronisation, identifiers, integrity, access, retention, rotation, export and backup. Test whether a reviewer can reconstruct the critical decision.
- A log becomes an accountability control when someone monitors it, anomalies trigger action and the record can support a post incident review.

[Sources]
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727
- https://www.nist.gov/itl/ai-risk-management-framework

Incident Response Closes the Accountability Loop

- Define how staff identify, contain, report, investigate and learn from an incident or near miss. Preserve evidence and protect affected people. Assign authority for suspension, external notification, communication and restart.

- Review technical, human, organisational and supply chain causes. Corrective actions need owners, deadlines and verification.

- Share lessons across similar systems and update the risk assessment.

[Sources]

- https://www.nist.gov/itl/ai-risk-management-framework
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727

An Assurance Case Links Safety Claims to Evidence

- An assurance case states the claim, the reasoning that supports it and the evidence that can be reviewed. Example: the robot remains within safe separation limits because sensing, speed control, protective stop and verification evidence work together.
- Record assumptions, operating conditions, confidence, missing evidence and counter evidence. Link each control to a measurable acceptance criterion.
- The case must identify its system version and approval date. It supports review and cannot replace testing or independent challenge.

[Sources]

- https://www.nist.gov/itl/ai-risk-management-framework
- https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2

Change Control Prevents Assurance Evidence from Becoming Obsolete

- Baseline the approved model, software, hardware, data, configuration and operating conditions. Classify every change and assess its effect on hazards, interfaces, human work and previous evidence.
- Require targeted regression tests, independent review for safety relevant changes, staged release, monitoring and rollback. Notify actors who rely on the changed evidence.
- A substantial modification can change regulatory responsibilities. Check the AI Act and applicable product or sector rules before release.

[Sources]

- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727
- https://unece.org/transport/documents/2021/03/standards/un-regulation-no-156-software-update-and-software-update

03

Learn Accountability Through Cases

Analyse project and official cases across technical, human and organisational layers.

Project Case: AI Self-Driving Taxis

- The Ethical Engineer case examines self-driving taxis as a service that may improve mobility and efficiency while raising safety, accountability, employment and public trust questions.

- Engineering focus: operating domain, perception limits, fallback, remote support, passenger information, incident ownership and social impact.

- Open the Ethical Engineer self-driving taxi case. Identify who can authorise operation, who can stop the fleet and which evidence they need.

[Sources]

- https://ethicalengineer.eu/download/223/
- https://ethicalengineer.eu/ethical-engineer-case-studies/

Official Case: Uber Automated-driving Crash

- In 2018 an Uber test vehicle struck and killed a pedestrian in Tempe, Arizona. The NTSB identified inadequate safety risk assessment, ineffective oversight and weak safety culture. Collision warning and automatic emergency braking functions had been dis.
- The case shows that human monitoring cannot compensate for poor system and organisational design.
- Trace the failure through sensing, classification, braking, operator role, supervision and governance.

[Sources] https://www.ntsb.gov/investigations/Pages/HWY18MH010.aspx

Apply a Layered Failure Analysis to the Uber Case

A strong bias audit turns evidence into an accountable argument

01	02	03	04	05
Technical Layer	**Human Layer**	**Operational Layer**	**Organisational Layer**	**Regulatory Layer**
Perception, prediction and disabled braking.	Vigilance demand, interface and response time.	Test procedure, driver monitoring and route control.	Safety risk management, staffing and oversight.	Test permissions and external assurance.

Use five whys carefully. Continue beyond operator error to the upstream decision, evidence gap and control that could have prevented or mitigated the event.

[Sources] ttps://www.ntsb.gov/investigations/Pages/HWY18MH010.aspx

Official Case: Boeing 737 MAX

- A United States House Committee investigation examined the design, development and certification of the Boeing 737 MAX after two crashes killed 346 people. The report considered MCAS design, sensor dependence, pilot information, certification and organisational pressures.
- Engineering focus: single point failure, authority of automation, response assumptions, change communication and independent review.
- Ask which records and decisions were required at each approval gate.

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

Accountability in the Boeing Case Crosses Organisations

- A component explanation is insufficient. Trace decisions across aircraft design, software, safety assessment, training, certification, management and airline operation.
- Identify who generated hazard evidence, who reviewed it, who controlled information and who approved the resulting system. Examine whether production and schedule pressures affected safety decisions.
- Build a traceability chain that makes assumptions, changes and dissent visible before certification.

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

Project Case: AI in Pulsatile Heart Pumps

- The Ethical Engineer case examines the tension between standardised medical device design and personalised care. Standardization facilitates research, evaluation, and regulatory approval, while individual patient needs may require different settings and monitoring.

- Engineering focus: intended population, physiological variation, safe limits, clinician authority, alarms, evidence and responsibility for adjustment.

- Open the pulsatile heart pump case. Define who can change a setting and who is accountable for its validation.

[Sources]

- https://ethicalengineer.eu/download/235/
- https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng

Safety Accountability Uses Domain Specific Evidence

01
Medical Devices

Risk management, clinical evidence, quality systems, vigilance and clinician authority.

02
Machinery & Robotics

Protective measures, safety functions, validation, instructions, maintenance and controlled modification.

03
Aviation & Vehicles

Certification or type approval, human factors, operational limits, software update control & independent assurance.

04
Critical Infrastructure

Continuity, cybersecurity, operator authority, incident coordination and recovery.

The accountability map can be reused. The acceptance evidence and competent authorities differ by domain.

[Sources] https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng.
https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727
https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2
https://unece.org/transport/documents/2021/03/standards/un-regulation-no-156-software-update-and-software-update

Compare Accountability Across the Three Cases

- The taxi case emphasises operating domain and fleet governance. The Uber investigation adds safety culture and test oversight. The Boeing case shows certification, information and organisational pressure. The heart pump case highlights clinical authority and individual variation.
- For each, identify the critical decision, accountable owner, required evidence, stop authority and remedy.
- Separate reusable accountability controls from the domain evidence required for release.

[Sources]

- https://ethicalengineer.eu/download/223/
- https://www.ntsb.gov/investigations/Pages/HWY18MH010.aspx
- https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation
- https://ethicalengineer.eu/download/235/

Applied Task: Conduct a Structured Failure Analysis

Choose Uber, Boeing or the heart pump case.

1. Define the unwanted event and system boundary.
2. Build a timeline of decisions and conditions.
3. Identify technical, human, operational and organisational contributors.
4. Locate missing or ignored evidence.
5. Propose prevention, detection, mitigation and recovery controls.
6. Assign owners and verification evidence.

Output:

One page failure map and a two minute accountability statement.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://www.ntsb.gov/investigations/Pages/HWY18MH010.aspx
- https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation
- https://ethicalengineer.eu/download/235/

Dilemma: Release After an Unexplained Near Miss

- A supplier reports that a safety relevant update passes the standard benchmark. A representative field test then produces one unexplained near miss. Management wants release to meet a contractual date.
- Decide GO, REVISE or STOP. State which evidence is missing, who can delay release and whether a bounded test is defensible. Define its operating limits, monitoring, stop trigger and rollback.
- Record the strongest dissenting view and the evidence that would change the decision.

[Sources]
- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://www.nist.gov/itl/ai-risk-management-framework

04

Produce a Traceable Engineering Decision

Complete an accountability record, release review, learning check and dossier contribution.

Project Case A: Autonomous Warehouse Robot

- A mobile robot moves pallets through a mixed warehouse. A software update improves route efficiency but increases close passes near workers. The vendor supplies model metrics but no system level safety test. Local managers want immediate deployment before a seasonal peak.
- Decide GO, REVISE or STOP. Define the operating boundary, detection, speed and separation controls, update approval, logs, stop authority and restart evidence.
- Name the accountable owner for each decision.

[Sources]
- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727

Completed Accountability Record: Warehouse Robot Update

Boundary

Mixed pedestrian aisle 3, marked crossings, dry floor, supervised day shift and maximum speed of 1.2 m/s.

Decision Revise

Full deployment is withheld. A bounded off peak pilot may proceed.

Owners

Operations director owns release. Safety engineer owns separation controls. Supplier owns update evidence. Shift supervisor has stop authority.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727

Completed Accountability Record: Warehouse Robot Update

Evidence

Version record, hazard review, 200 route trials, minimum separation results, protective stop test, worker walkthrough and rollback test.

Controls

Speed limit, geofenced crossings, audible approach signal, independent protective scanner, incident log and immediate rollback.

Review

Approve only if every protective test passes, no unexplained close pass occurs and worker feedback is resolved. Reassess after 50 operating hours.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727

Project Case B: Medical Alert Prioritisation

- An AI service ranks patient alerts for a remote monitoring team. A model update reduces false alarms overall, yet its effect on rare cardiac conditions is uncertain. Staff can see the score but cannot identify the model version or reason for the ranking.
- Define the validation, fallback, interface, logging, escalation and incident controls required before release.
- State who accepts residual risk and which result triggers immediate rollback.

[Sources]
- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng

Team Activity: Build the Accountability Chain

Choose the warehouse robot or medical alert case. Assign system owner, safety lead, supplier, operator, reviewer and recorder roles.

1. Define the critical decision and system boundary.
2. Map requirement, implementation, verification, approval, operation and incident response.
3. Add the evidence owner, control owner and stop authority at each stage.
4. Agree a GO, REVISE or STOP conclusion.
5. Produce a two minute defence based on the traceability chain.

[Sources]. Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.

A strong accountability analysis connects authority with evidence

- A strong analysis defines the system boundary and unwanted event. It distinguishes assigned duty, accountable decision and legal context. The failure analysis includes technical, human and organisational contributors.
- Controls address prevention, detection, mitigation and recovery. Every control has an acceptance criterion and owner. The record preserves assumptions, dissent, residual risk, monitoring and restart conditions.
- Alternative decisions can be valid when their evidence, authority and review conditions are explicit.

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

Worksheet: Accountability & Traceability Record

System, intended use and operating boundary: .....

Critical decision and unwanted event: .....

Responsible roles and accountable decision owner: .....

Technical, human and organisational contributors: .....

Safety controls and acceptance tests: .....

Assurance evidence, logs and change records: .....

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.

Continues On Next Slide

Decision: GO REVISE STOP .....

Residual risk, stop authority and review trigger: .....

Incident, remedy and restart route: .....

Dissent or alternative view: .....

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.

Role Play: Release Review Under Schedule Pressure

- A supplier update passes the benchmark, while a field test reveals one unexplained near miss. Management wants release to meet a contractual date.
- Roles: engineering manager, safety engineer, product owner, supplier, operator representative and independent reviewer.
- Decide which evidence is required, who can delay release, how dissent is recorded and which conditions permit a limited test. State one unacceptable argument and one valid basis for approval.

[Sources]
- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- ttps://www.nist.gov/itl/ai-risk-management-framework

Individual Reflection & Transfer

Write a short engineering argument.

1. Which decision required clear accountability?
2. Which control failed or remained absent?
3. Which evidence should have changed the decision?
4. Which evidence should have changed the decision?
5. Could the team identify, explain and correct the failure?
6. How would you design accountability into your next project?

Output:

Transfer the reasoning to another discipline and state one change in roles, timing or evidence.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.

Quiz

1. How does accountability differ from responsibility?
2. Why can two sensors fail as one control?
3. What makes human oversight meaningful?
4. What belongs in an assurance case?
5. Which evidence supports incident reconstruction?
6. Why should a safety relevant update trigger renewed review?
7. Who should accept residual risk?

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.

Quiz Answers & Teaching Rationale

1. Accountability adds justification, evidence, review and correction.
2. Sensors may share a common cause, data or dependency.
3. Meaningful oversight needs authority, information, competence and response time.
4. A claim, supporting argument, evidence, assumptions, limits and version.
5. Versioned logs, configuration, inputs, outputs, alarms, actions and changes.
6. The change may invalidate hazard assumptions, controls or earlier evidence.
7. A named role with authority, competence and organisational mandate.

[Sources]. Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning. https://www.nist.gov/itl/ai-risk-management-framework

Learning Outcome Check

Before finishing, confirm that you can:

1. Distinguish responsibility, accountability and liability.
2. Map decision and control ownership across a lifecycle.
3. Analyse layered failure causes.
4. Specify resilience, fallback, handover, logging, assurance and change controls.
5. Produce a traceable decision with residual risk and review.

Evidence:

Responsibility map, failure analysis, assurance argument and accountability record.

[Sources] Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.

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

Thank You

You have completed **Module 4**

www.ethicalengineer.eu

Co-funded by the European Union