

Module 5

Values Based Design in AI Engineering



Co-funded by
the European Union



Ethical Engineer

www.ethicalengineer.eu

CONTENTS

- 01 Discover Stakeholders & Values
- 02 Translate Values into Design
- 03 Compare Values Across Cases
- 04 Defend a Values-based 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	Energy, water and hardware Societal and lifecycle impacts Metrics, governance and redesign

Module Orientation

This module turns stakeholder values into requirements, design alternatives, test evidence and a defensible engineering decision.

Purpose

Equip engineering students to investigate values, translate them into design criteria and defend choices with stakeholder and technical evidence.

Audience

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

Builds On

Modules 1 to 4: ethics, transparency, fairness, privacy, responsibility, safety and traceable engineering decisions.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>

Module Orientation

This module turns stakeholder values into requirements, design alternatives, test evidence and a defensible engineering decision.

Learning Sequence

Discover stakeholder values, translate them into design, compare cases and defend a values-based decision.

Learner Outputs

Stakeholder map, value hierarchy, requirements, option comparison and values-based design record.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>

Learning Outcome

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

1. Identify direct and indirect stakeholders, power differences and relevant values.
2. Conduct conceptual, empirical and technical Value Sensitive Design investigations.
3. Translate values into requirements, acceptance criteria and traceable design evidence.
4. Compare data, model, interface and system options using thresholds, evidence and uncertainty.
5. Produce a values-based design record with safeguards, residual tensions and review conditions.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>
- <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Values are already present in Engineering Choices

- Objectives define success. Datasets decide whose experience is visible. Thresholds distribute risk. Interfaces shape attention and control.
- AI engineering is socio-technical: behaviour emerges from people, data, models, interfaces, organisations and operating context.
- Values-based design makes these choices explicit, evidence-based and reviewable. The engineering task is to show how a value changed a decision and how the requirement will be verified.

[Sources]

- Original Module 5 deck supplied by the Ethical Engineer project consortium.
- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>

Frameworks & Methods Guide Engineering Decisions

Value Sensitive Design uses conceptual, empirical and technical investigations to examine human values throughout design. IEEE 7000 provides a systems-engineering process for eliciting values and tracing them into concepts, requirements and ethical risk treatment.

The EU Ethics Guidelines identify requirements for trustworthy AI. The EU AI Act creates legal obligations for defined systems and actors. These sources guide related work and do not replace stakeholder evidence or domain engineering standards.

Use the current legal text and record which method, obligation and evidence support each significant decision.

[Sources]

- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>
- <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 Duties & Engineering Accountability

The EU AI Act assigns duties by system, risk category and actor role. Engineering teams must connect these duties with domain law, safety standards and organisational controls.

For each significant decision, record: decision owner, evidence owner, reviewer, control authority and stop authority.

Accountability becomes practical when responsibility, evidence, approval and escalation routes are visible before deployment.

[Sources]

- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>
- <https://www.nist.gov/itl/ai-risk-management-framework>
- Latest Module 5 deck supplied by the Ethical Engineer project consortium.

01

Discover Stakeholders & Values

Define values in context,
examine power and
gather evidence from
affected stakeholders.



Values-Based Design Connects Ethical Purpose with System Choices

A value expresses what people or institutions consider important, such as safety, dignity, autonomy, justice, privacy, accessibility, reliability or sustainability. Its meaning depends on the people, practice and operating context.

Values influence objectives, data, functions, interfaces, defaults, business rules and maintenance. They can support one another or create tension.

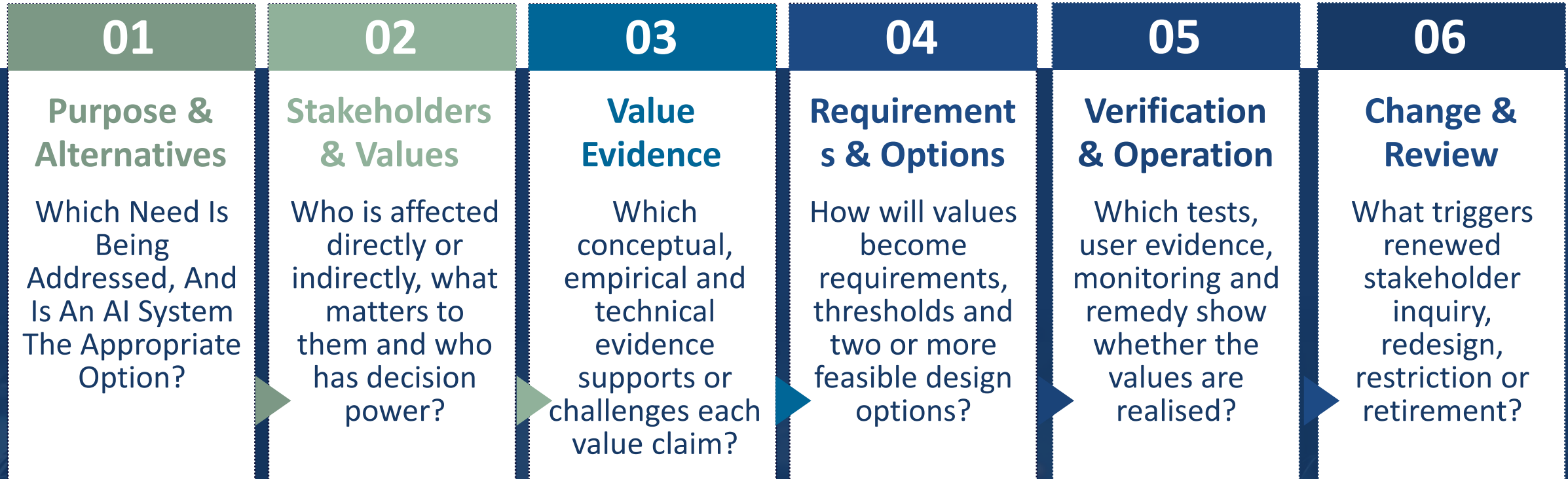
A values-based decision identifies stakeholders, evidence and alternatives, then traces the selected value interpretation into requirements, tests, owners and review conditions.

[Sources]

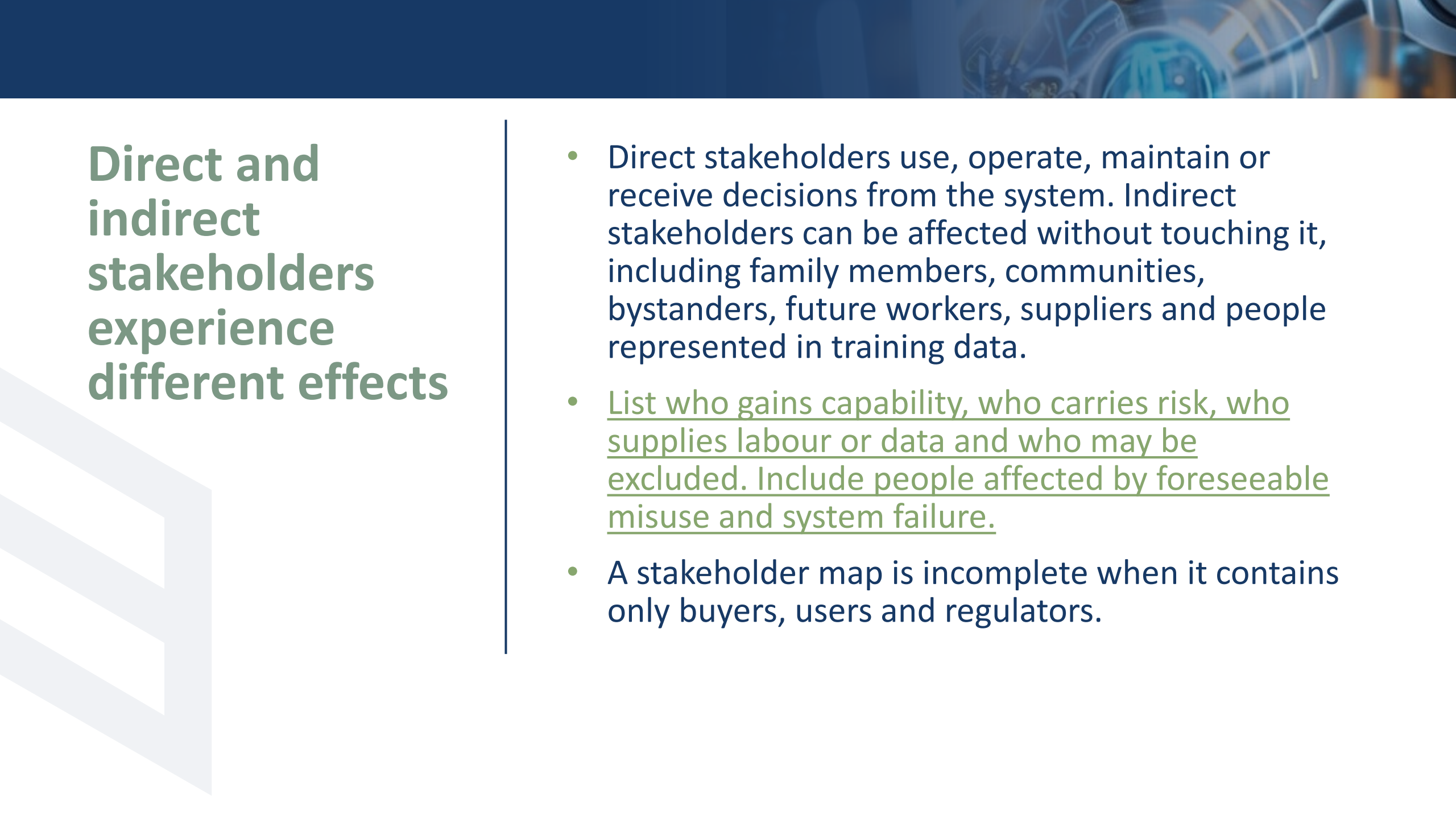
- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>

Map Values Across the Engineering Lifecycle

Choose one engineering decision. At each stage, identify stakeholders, value claims, evidence, design response and review.

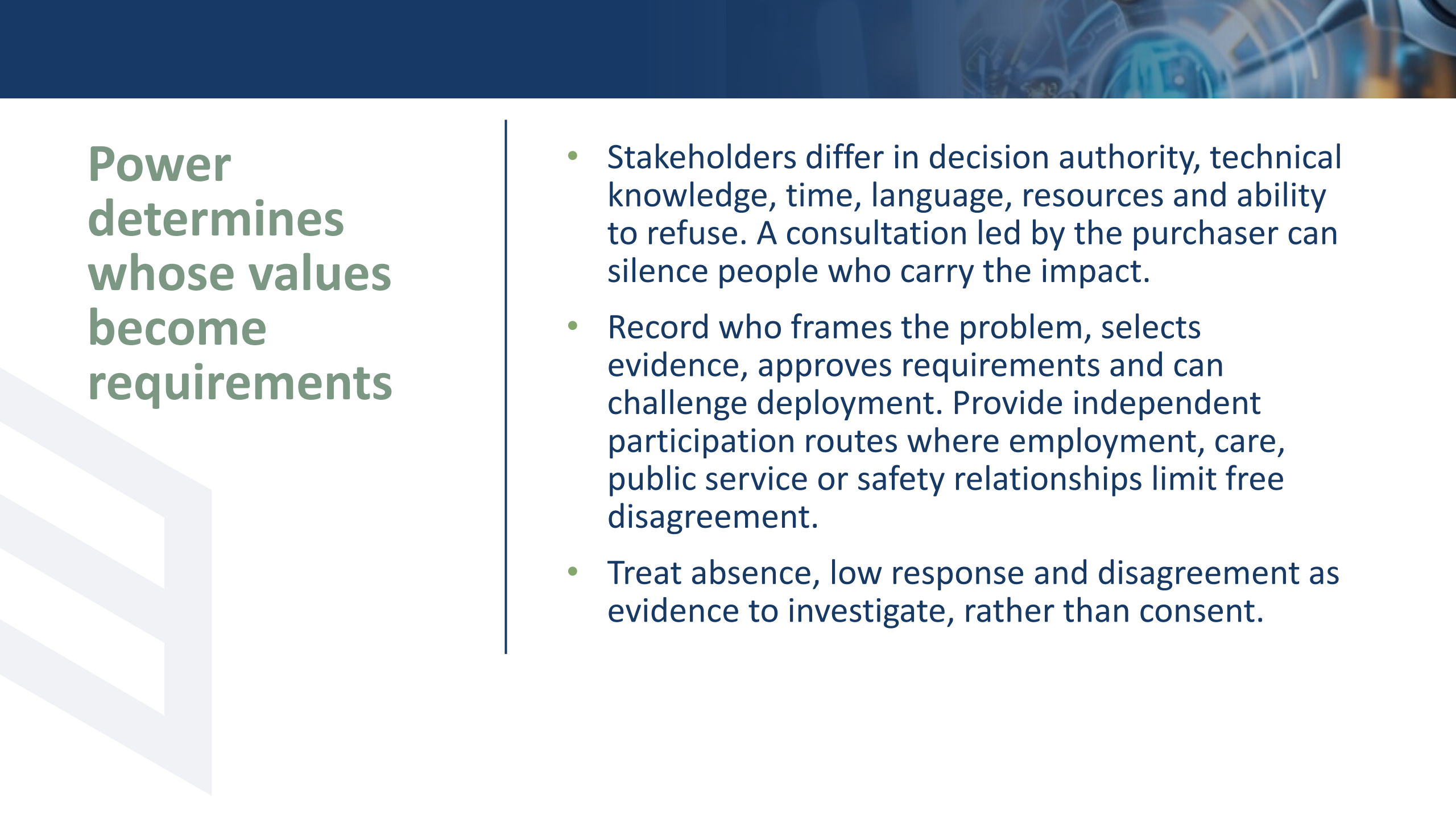


OUTPUT A values map linking each stage to stakeholders, evidence, a design response, acceptance criteria and review trigger.



Direct and indirect stakeholders experience different effects

- Direct stakeholders use, operate, maintain or receive decisions from the system. Indirect stakeholders can be affected without touching it, including family members, communities, bystanders, future workers, suppliers and people represented in training data.
- List who gains capability, who carries risk, who supplies labour or data and who may be excluded. Include people affected by foreseeable misuse and system failure.
- A stakeholder map is incomplete when it contains only buyers, users and regulators.



Power determines whose values become requirements

- Stakeholders differ in decision authority, technical knowledge, time, language, resources and ability to refuse. A consultation led by the purchaser can silence people who carry the impact.
- Record who frames the problem, selects evidence, approves requirements and can challenge deployment. Provide independent participation routes where employment, care, public service or safety relationships limit free disagreement.
- Treat absence, low response and disagreement as evidence to investigate, rather than consent.



Define each value in the specific engineering context

MSafety may mean preventing injury, maintaining a safe state or supporting timely intervention. Autonomy may concern informed choice, control, freedom from manipulation or the practical ability to refuse. Fairness may concern outcomes, process, access or remedy.

Write a contextual definition, affected stakeholders, a plausible threat and observable evidence. Avoid using a broad value word as if everyone shares one interpretation.

The definition should be precise enough to guide a design option and an acceptance test.



Stakeholder evidence must change design choices

Use interviews, observation, co-design, surveys, usability studies, complaints and field evidence according to the decision.

Include affected people, end users and technical, legal, ethical and domain experts. Experts do not replace those who carry the impact.

Record what changed: the problem definition, requirement, option, test or review condition — and who remains unrepresented.

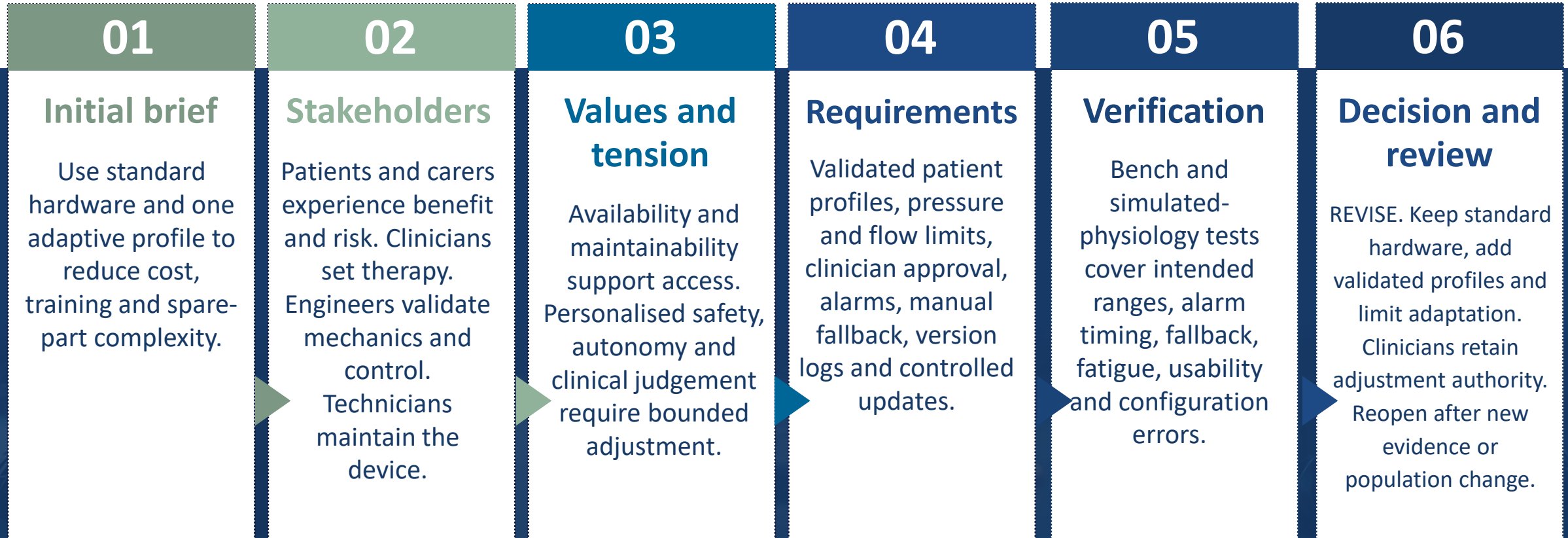


Value tensions require evidence and constraints

- Values may support one another: accessible explanations can improve autonomy, safety and trust. They can also conflict: detailed logging may support accountability while increasing privacy risk; personalisation may improve fit while reducing standardisation and maintainability.
- First define non-negotiable legal, safety or rights thresholds. Then compare feasible options using relevant evidence. Do not describe avoidable harm as a balanced trade-off.
- Record the residual tension, affected stakeholders and the evidence that would reopen the decision.

WORKED MECHANICAL ENGINEERING EXAMPLE: HEART PUMP DESIGN

A team proposes one AI-supported control profile for a pulsatile heart pump to simplify manufacture and maintenance. Stakeholder evidence shows that different patients need validated operating ranges and clear clinical authority.



VALUES CHANGED THE DESIGN A standardisation brief became a configurable mechanical and control architecture with patient-specific limits, clinical authority and reviewable evidence.

02

Translate Values Into Design

Create value hierarchies, requirements, tests, data and model interventions, documentation and lifecycle checkpoints.



Three Investigations Form One Iterative Design Method

Value Sensitive Design combines conceptual, empirical and technical investigations. Each can change the others as the problem, evidence and design develop.

Use the three investigations repeatedly. A stakeholder finding can change a requirement; a prototype can reveal a new stakeholder or value tension.

See the next slide...

[Sources]

- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>

Three Investigations Form One Iterative Design Method

01

Conceptual

Identify direct and indirect stakeholders, define values in context and examine tensions, rights and assumptions.

02

Empirical

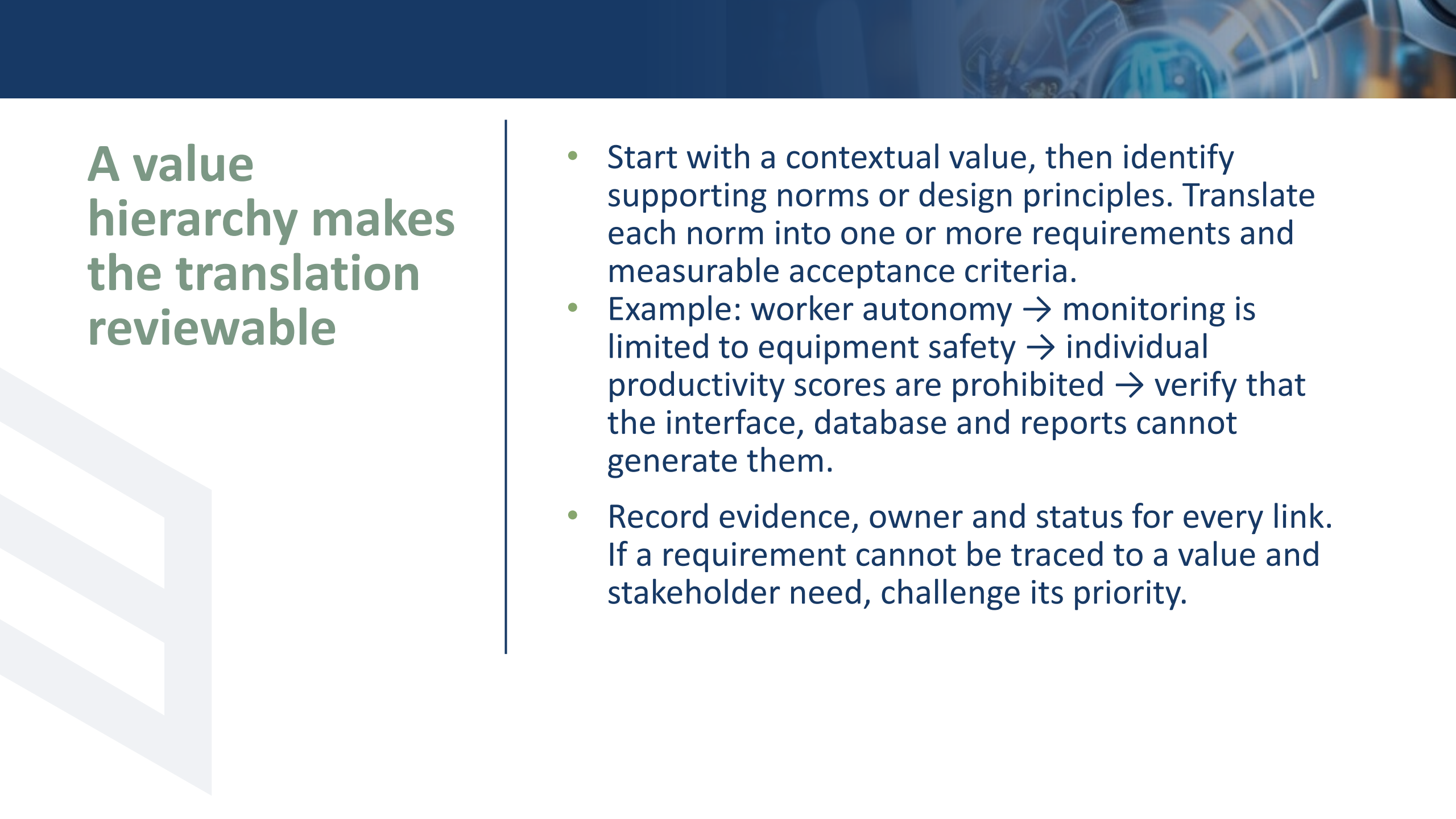
Investigate how people understand, experience and act around the technology using suitable social and field methods.

03

Technical

Analyse how existing and alternative designs support or undermine values, then prototype and verify requirements.

The output is a traceable chain: value definition, stakeholder evidence, design requirement, technical response, acceptance test, owner and review trigger.



A value hierarchy makes the translation reviewable

- Start with a contextual value, then identify supporting norms or design principles. Translate each norm into one or more requirements and measurable acceptance criteria.
- Example: worker autonomy → monitoring is limited to equipment safety → individual productivity scores are prohibited → verify that the interface, database and reports cannot generate them.
- Record evidence, owner and status for every link. If a requirement cannot be traced to a value and stakeholder need, challenge its priority.

Ethical user stories need testable acceptance criteria

- A useful story identifies the affected role, capability or protection and the reason it matters. It also states misuse, exclusion or failure conditions.
- Example: As a maintenance technician, I need the system to show sensor confidence and recent configuration changes so that I can judge whether an alert is actionable.
- Acceptance evidence: representative task test, readable uncertainty, version history, alarm response time and a documented fallback. Link the story to the requirement and verification record.

Data interventions carry value choices

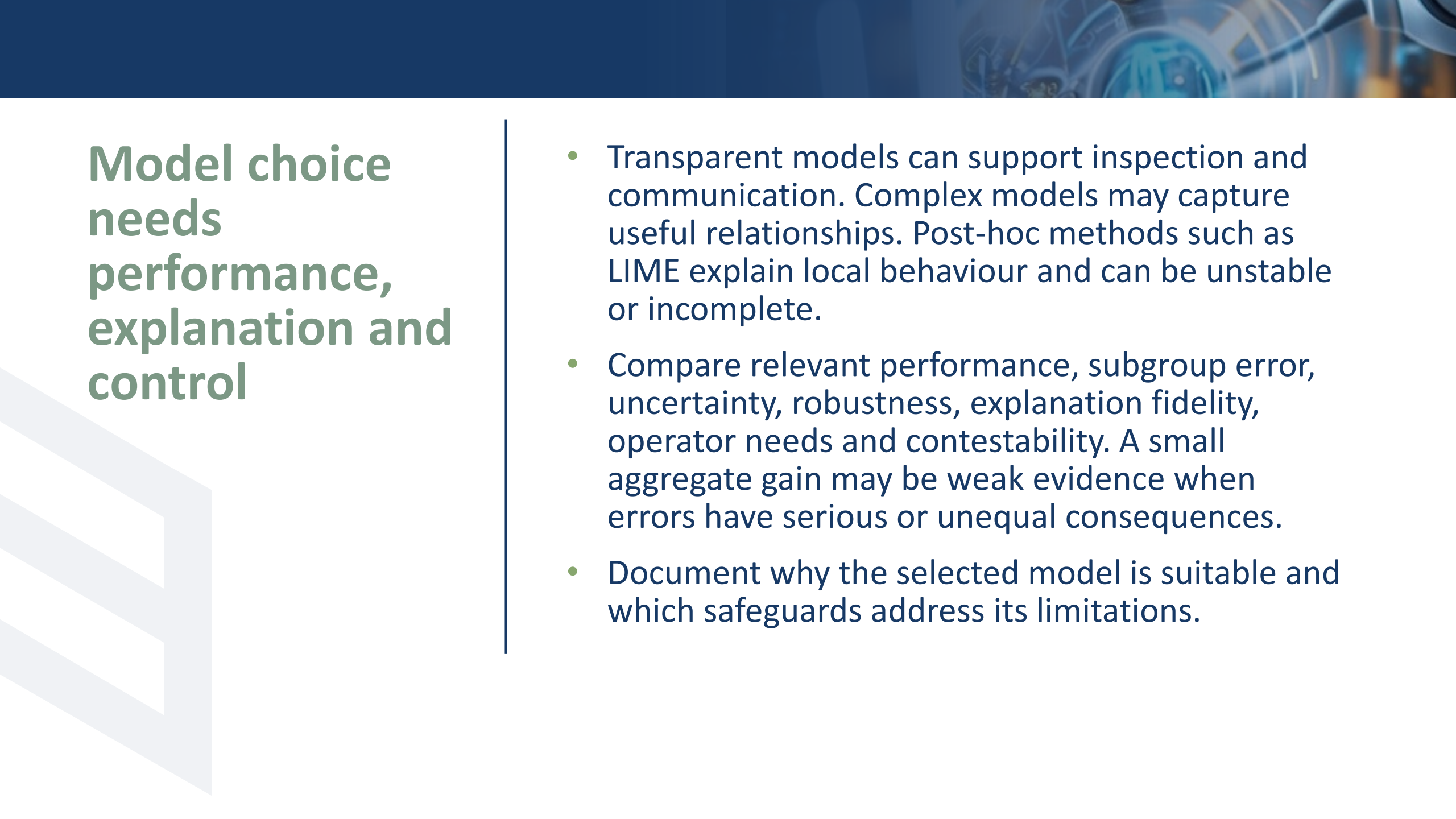
- Sampling, labels, proxies, missing data rules and augmentation determine which cases receive attention. SMOTE or class weighting can support rare classes while changing calibration. Transfer learning imports assumptions from another domain. PCA or feature selection can improve tractability while removing information relevant to a subgroup.
- For each intervention, state the intended value, affected group, mechanism, comparison baseline and evaluation evidence.
- Document provenance, limits, prohibited uses and the decision to retain or remove a feature.

Data imbalance and SMOTE require careful evaluation

Rare classes can reflect sampling, labelling, access or measurement problems. Investigate the cause before selecting a technical response.

Options include new data collection, stratified evaluation, class weighting and synthetic minority oversampling. SMOTE creates synthetic minority examples between observed cases; it can support learning while changing calibration or producing implausible cases when the feature space is unsuitable.

Compare the baseline and intervention using subgroup error, calibration, uncertainty and domain review. Record which value the intervention supports and which risks remain.



Model choice needs performance, explanation and control

- Transparent models can support inspection and communication. Complex models may capture useful relationships. Post-hoc methods such as LIME explain local behaviour and can be unstable or incomplete.
- Compare relevant performance, subgroup error, uncertainty, robustness, explanation fidelity, operator needs and contestability. A small aggregate gain may be weak evidence when errors have serious or unequal consequences.
- Document why the selected model is suitable and which safeguards address its limitations.



Interpretability and LIME provide local evidence

LIME fits a simple local surrogate around one prediction to estimate which features influenced that local result.

It can help engineers explore behaviour, discuss a case and identify questions for further testing. The explanation can change with sampling, parameters and correlated features.

LIME does not establish causality, global behaviour or fairness. Test stability across repeated runs and nearby cases, compare with domain knowledge and record the decision the explanation supports and the limits communicated to users.



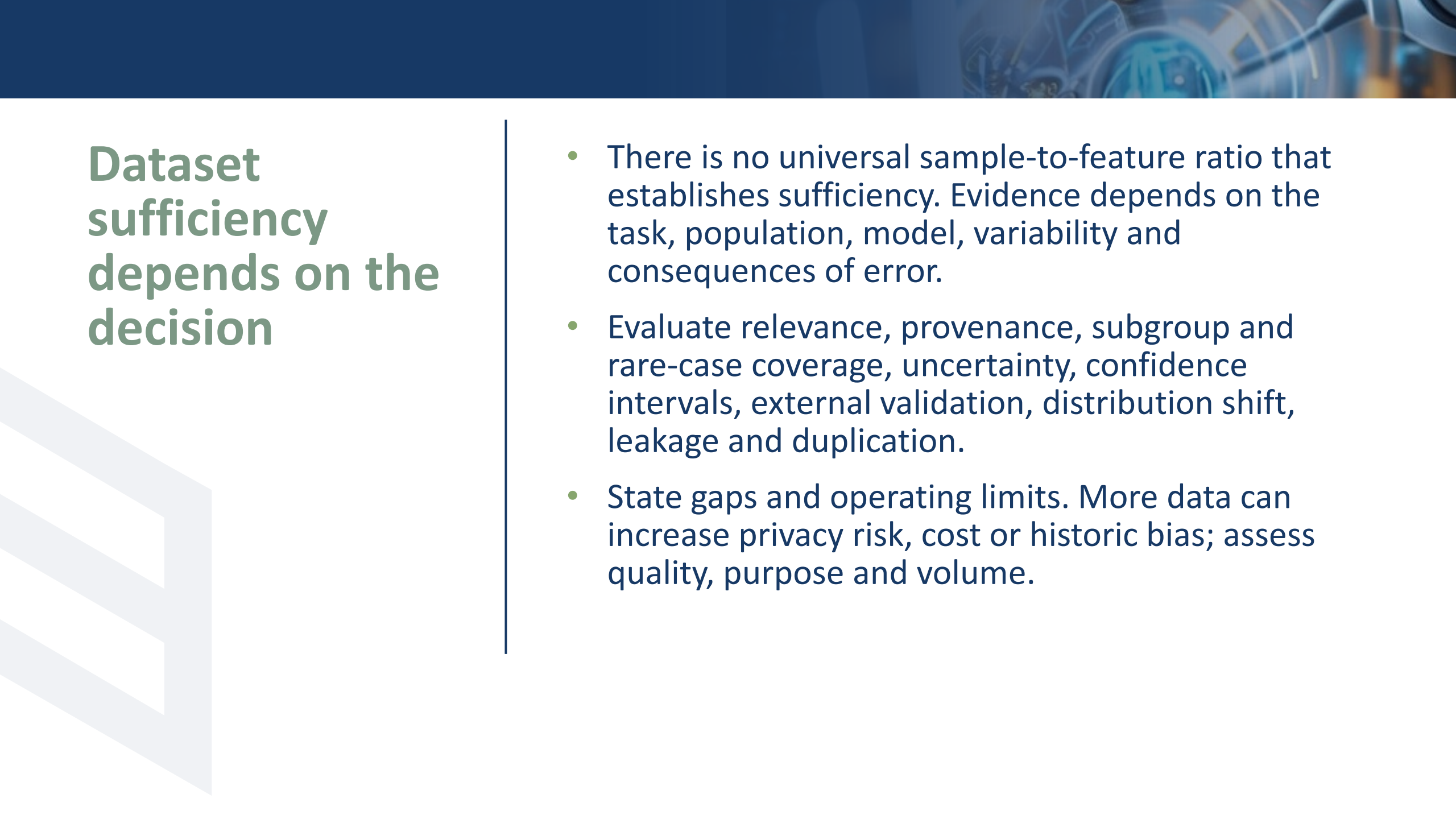
Interfaces shape attention, authority and challenge

Defaults, alert order, confidence display, labels and workflow timing influence reliance.

Automation changes the human role. When operators become supervisors, reliable automation can reduce vigilance and slow intervention.

Test whether users can recognise limits, regain situational awareness, act in time, override, seek review and recover after error.

Record the information each audience needs and the action the interface must support.



Dataset sufficiency depends on the decision

- There is no universal sample-to-feature ratio that establishes sufficiency. Evidence depends on the task, population, model, variability and consequences of error.
- Evaluate relevance, provenance, subgroup and rare-case coverage, uncertainty, confidence intervals, external validation, distribution shift, leakage and duplication.
- State gaps and operating limits. More data can increase privacy risk, cost or historic bias; assess quality, purpose and volume.

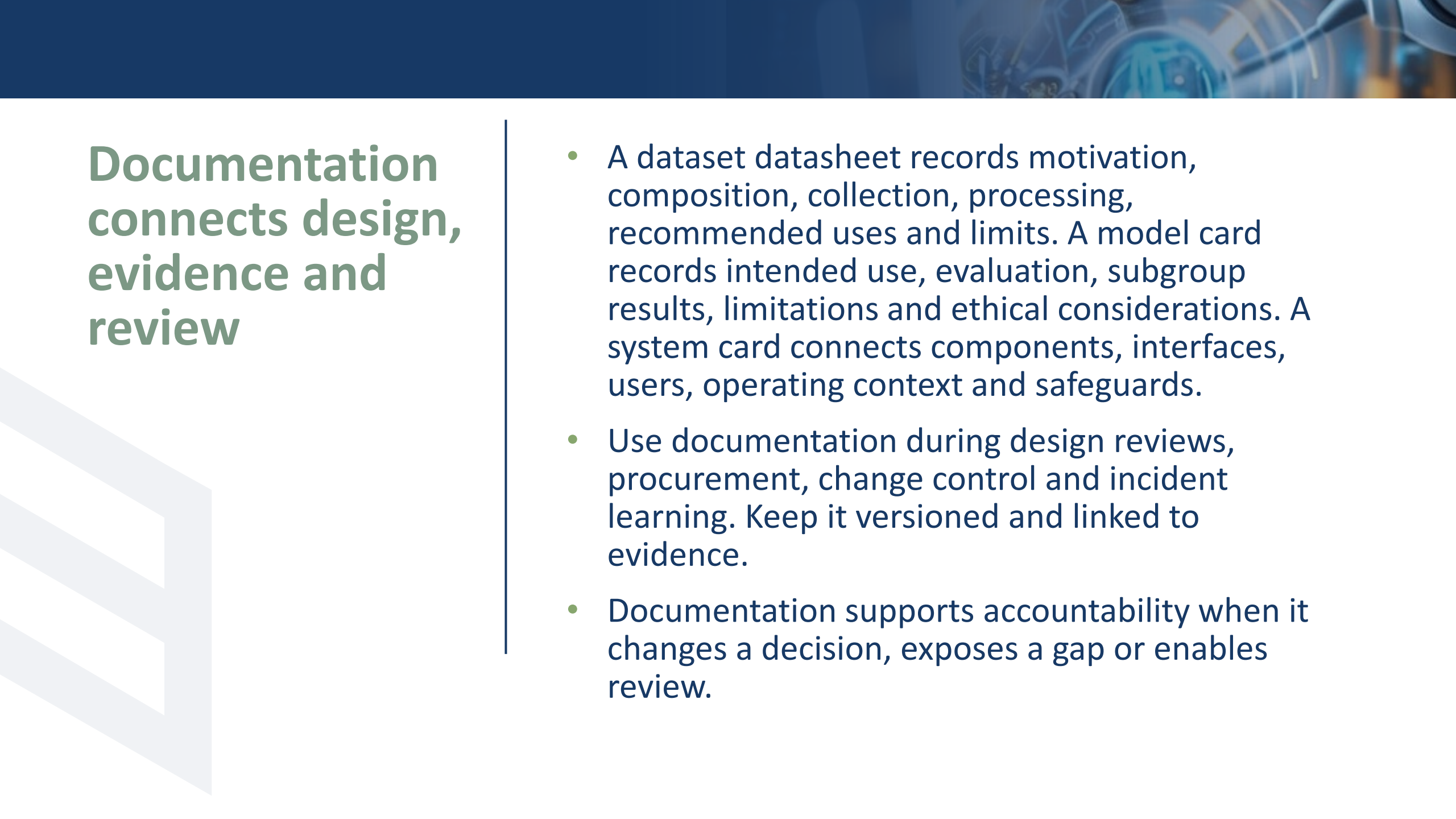


Data volume, dimensionality and transfer learning

There is no universal sample rule. Sufficiency depends on the task, population, variability, model and consequences of error.

High-dimensional data can create sparsity and overfitting even when the total sample appears large. Use learning curves, feature review, regularisation, uncertainty estimates and external validation.

Transfer learning and augmentation can reduce data demands while importing assumptions from another domain. Check domain fit, subgroup coverage and distribution shift.



Documentation connects design, evidence and review

- A dataset datasheet records motivation, composition, collection, processing, recommended uses and limits. A model card records intended use, evaluation, subgroup results, limitations and ethical considerations. A system card connects components, interfaces, users, operating context and safeguards.
- Use documentation during design reviews, procurement, change control and incident learning. Keep it versioned and linked to evidence.
- Documentation supports accountability when it changes a decision, exposes a gap or enables review.

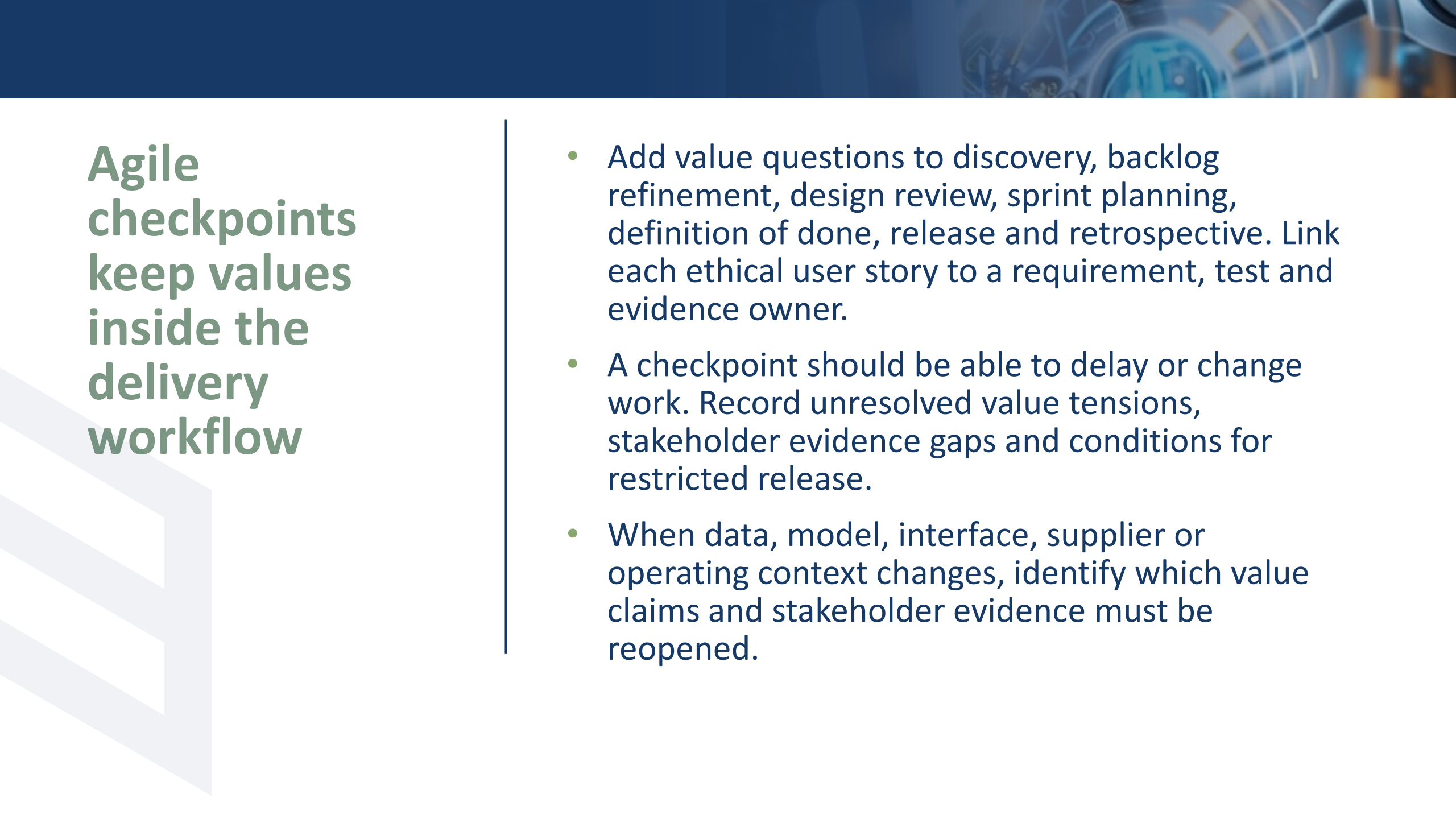


Datasheets, model cards and system cards

Dataset datasheets record motivation, composition, collection, processing, intended uses and limits.

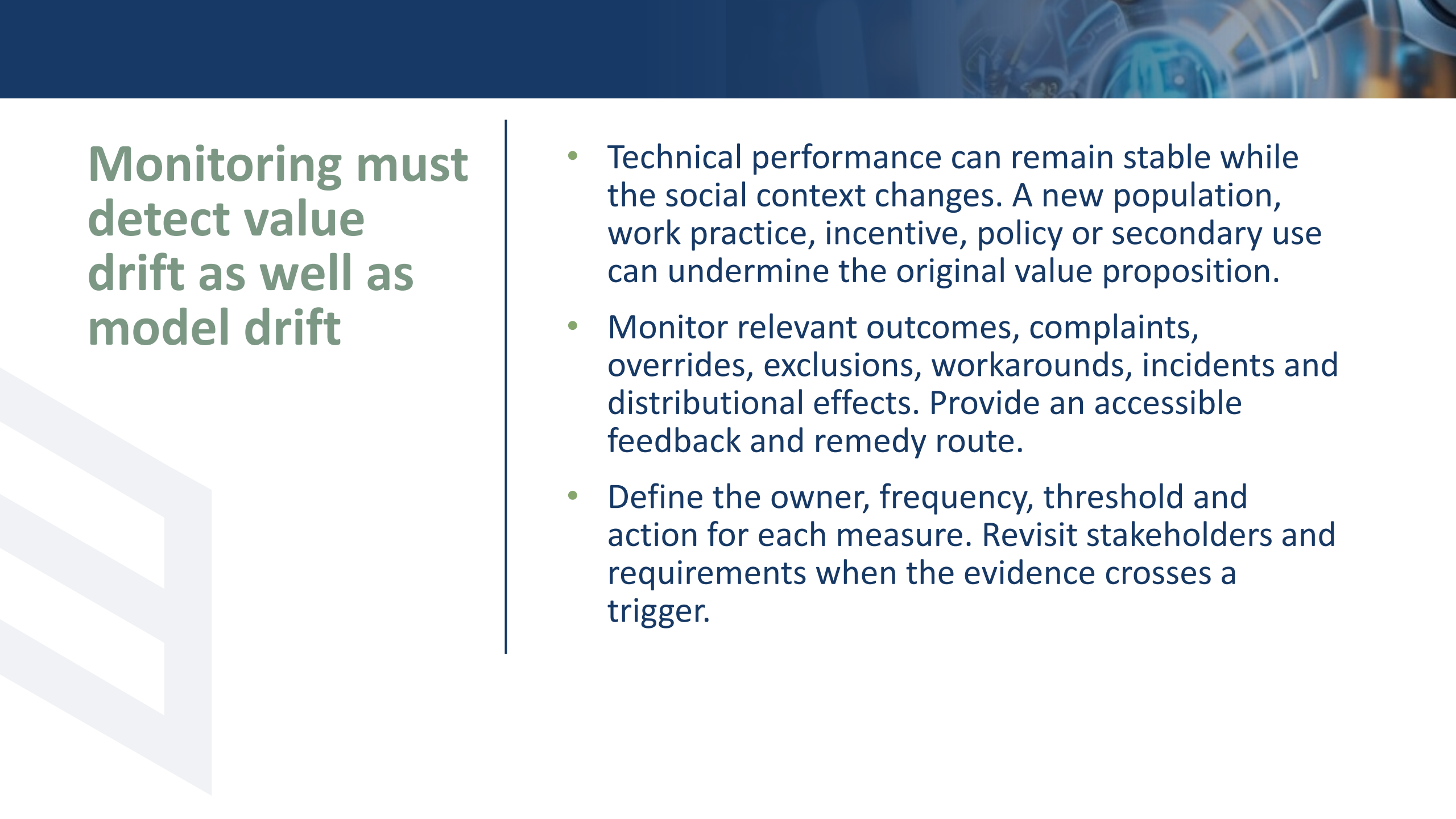
Model cards record intended use, evaluation, subgroup results, limitations and ethical considerations. System cards connect components, interfaces, users, context and safeguards.

Keep records versioned and linked to evidence, approvals and change control. Where the EU AI Act applies, connect them with technical documentation, risk management and post-market monitoring evidence.



Agile checkpoints keep values inside the delivery workflow

- Add value questions to discovery, backlog refinement, design review, sprint planning, definition of done, release and retrospective. Link each ethical user story to a requirement, test and evidence owner.
- A checkpoint should be able to delay or change work. Record unresolved value tensions, stakeholder evidence gaps and conditions for restricted release.
- When data, model, interface, supplier or operating context changes, identify which value claims and stakeholder evidence must be reopened.



Monitoring must detect value drift as well as model drift

- Technical performance can remain stable while the social context changes. A new population, work practice, incentive, policy or secondary use can undermine the original value proposition.
- Monitor relevant outcomes, complaints, overrides, exclusions, workarounds, incidents and distributional effects. Provide an accessible feedback and remedy route.
- Define the owner, frequency, threshold and action for each measure. Revisit stakeholders and requirements when the evidence crosses a trigger.



Apply mandatory thresholds before weighted option scoring

Begin with legal, safety, rights and professional constraints. Remove any option that fails a mandatory threshold unless the design can be revised.

Connect values-based design with risk management: identify the harm or value threat, assess evidence and consequence, choose a design control, verify the control and document residual risk and review triggers.

Then compare feasible options using criteria such as safety, autonomy, justice, privacy, explainability, performance, sustainability, cost and maintainability. A weighted total cannot cancel a serious rights or safety failure.

03

Compare Values Across Cases

Use project, engineering and clinical cases to examine context, proxies, inference, safety and trade-offs.



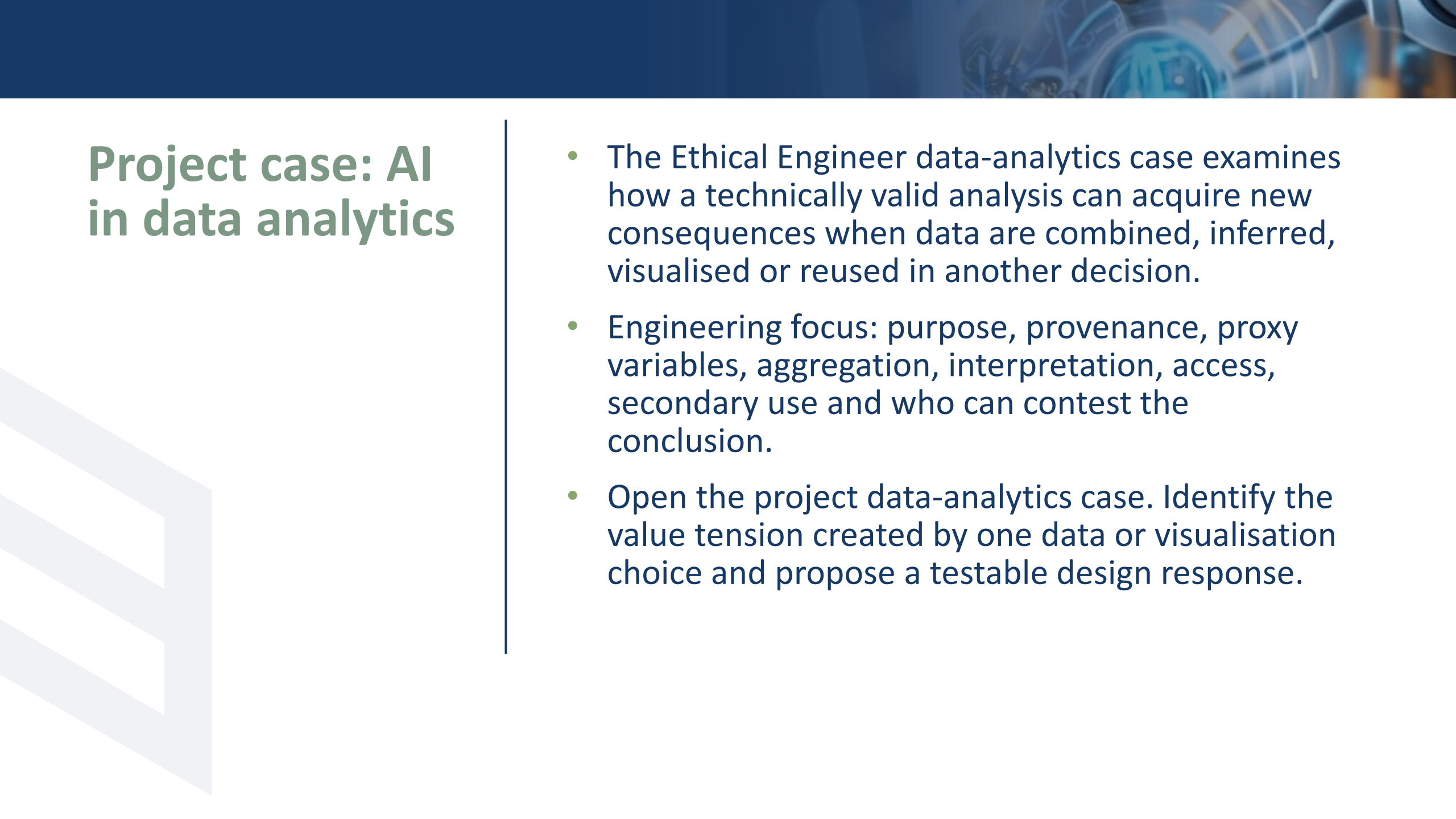


Project case: AI and hiring

- The Ethical Engineer hiring case examines how automation can shape access to work, reproduce historic patterns and shift decision power between applicants, recruiters and vendors.
- Engineering focus: purpose, evidence, relevant attributes, accessibility, explanation, recruiter authority, challenge and monitoring.
- Open the project hiring case. Identify one value claim made by the organisation and the evidence needed to test it.

Project case: assessing historic photographs

- The Ethical Engineer case explores AI-supported assessment of historic photographs. Cultural meaning, provenance and uncertainty can be reduced when a model treats the image as detached data.
- Engineering focus: source context, communities represented, sensitive inference, expert interpretation, uncertainty, access and correction.
- Open the historic photograph case. Define which stakeholders should influence the intended use and which inferences should be prohibited or qualified.



Project case: AI in data analytics

- The Ethical Engineer data-analytics case examines how a technically valid analysis can acquire new consequences when data are combined, inferred, visualised or reused in another decision.
- Engineering focus: purpose, provenance, proxy variables, aggregation, interpretation, access, secondary use and who can contest the conclusion.
- Open the project data-analytics case. Identify the value tension created by one data or visualisation choice and propose a testable design response.

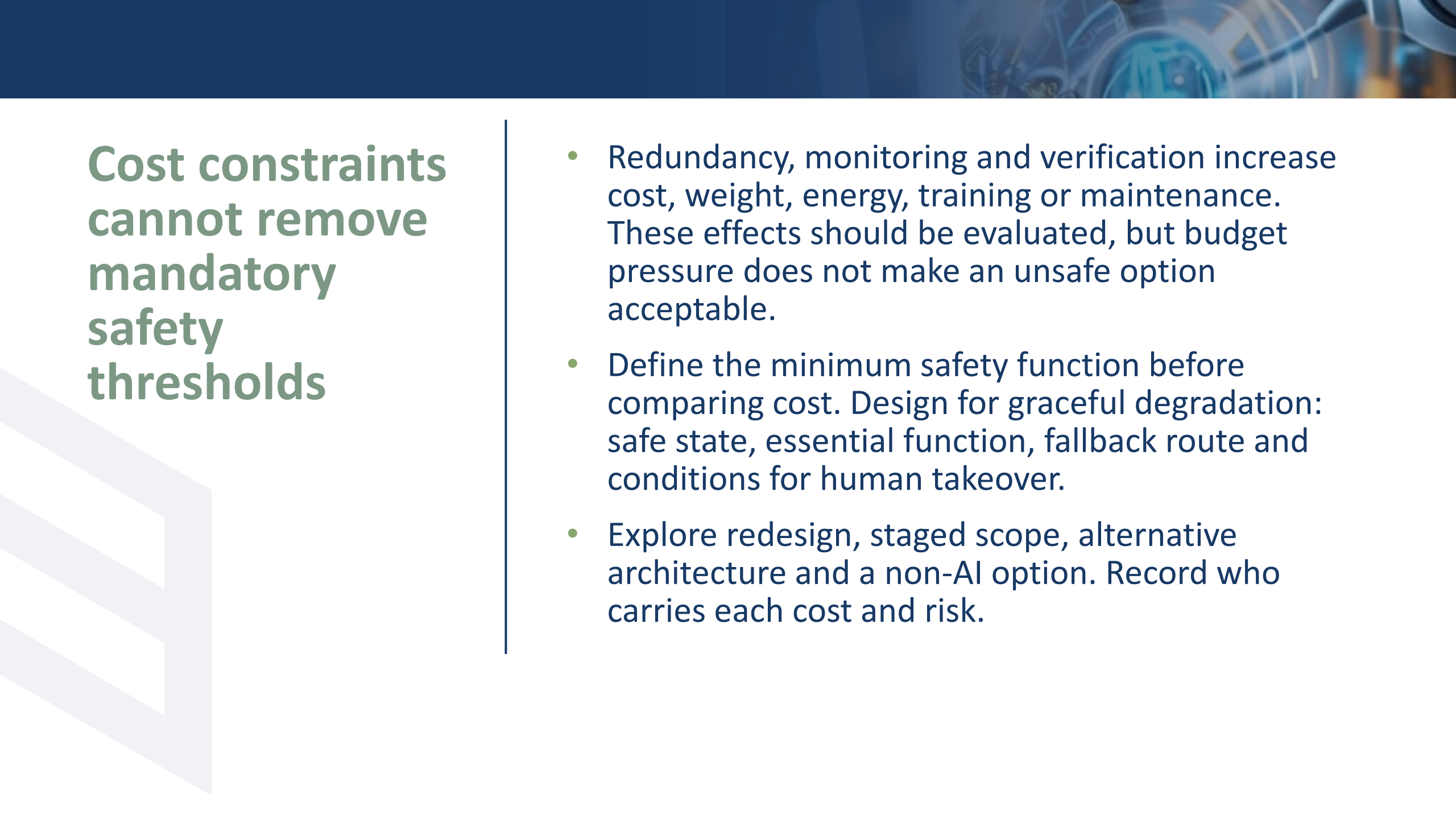
Project case: AI in pulsatile heart pumps

- The Ethical Engineer heart-pump case examines standardisation, access, personalised care and safety. A design that simplifies production and maintenance may fail to address physiological variation or clinical judgement.
- Engineering focus: intended population, patient values, safe mechanical and control limits, clinician authority, alarms, evidence and adjustment responsibility.
- Open the pulsatile heart pump case. Compare a fixed, configurable and adaptive design using mandatory safety thresholds.



The same method reveals different value tensions

- Hiring emphasises fair access, dignity, explanation and challenge. Historic photographs add cultural context, provenance and respectful interpretation. Data analytics highlights purpose, inference, proxy choices and secondary use. Heart pumps place safety, access, personalisation and clinical authority inside a mechanical system.
- For each case, define the engineering decision, stakeholders, contextual values, evidence gap, mandatory threshold and feasible options.
- Compare which methods transfer and which requirements depend on the domain.



Cost constraints cannot remove mandatory safety thresholds

- Redundancy, monitoring and verification increase cost, weight, energy, training or maintenance. These effects should be evaluated, but budget pressure does not make an unsafe option acceptable.
- Define the minimum safety function before comparing cost. Design for graceful degradation: safe state, essential function, fallback route and conditions for human takeover.
- Explore redesign, staged scope, alternative architecture and a non-AI option. Record who carries each cost and risk.



Failure cases expose the values embedded in upstream decisions

Uber shows how schedule, test governance and assumptions about human monitoring can shape safety. Boeing 737 MAX shows the consequences of sensor dependence, automation authority, information and organisational pressure. The clinical asthma case shows how a model can learn care pathways rather than physiology.

Locate the upstream design decision where better evidence, stakeholder challenge or a mandatory threshold could have changed the outcome.

Do not use procedure as a substitute for redesign when the hazard sits in the architecture. Failure analysis should produce a revised requirement and verification plan.



Clinical AI case: asthma and care pathway confounding

A hospital risk model appeared to treat asthma as protective for patients with pneumonia. In the training context, patients with asthma often received intensive care earlier. The model learned the care pathway and its outcomes rather than a safe physiological rule.

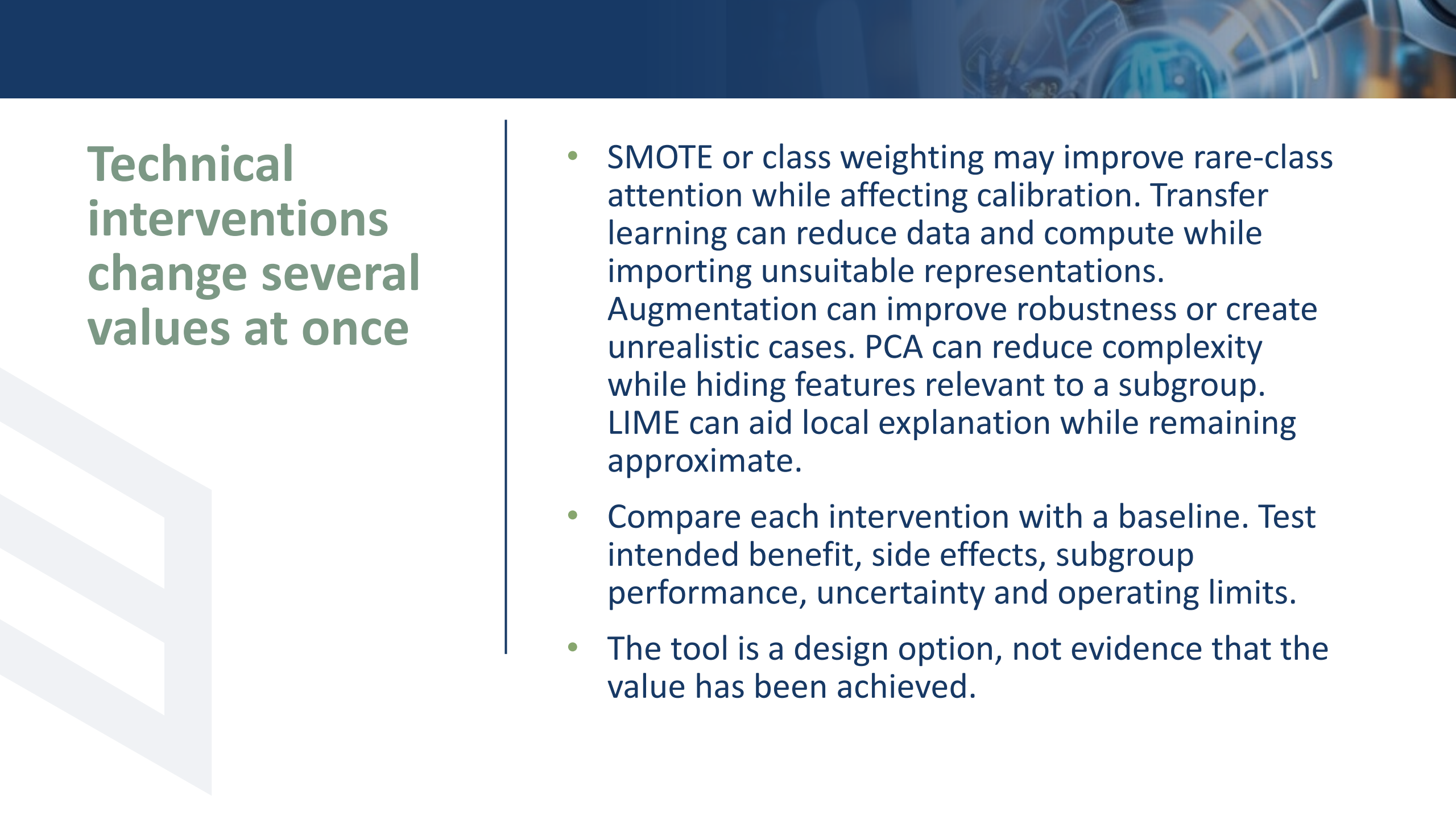
The pattern could become unsafe when used in a setting with different admission practices. Engineers should examine causal assumptions, care processes, transportability and domain shift.

Revise the requirement: high impact clinical recommendations require domain review, pathway analysis and validation in the intended setting before use.



Clinical AI case: health cost used as a proxy

A population health algorithm used predicted health cost as a proxy for health need. Unequal access and spending meant that cost did not represent illness equally across groups. People with similar health needs could receive different risk scores and access to support. Audit the objective, proxy, labels and allocation process. Compare outcomes by relevant groups and ask affected stakeholders whether the target reflects the intended value. Redesign the objective using direct measures of need, test the allocation effect and record the remaining uncertainty.



Technical interventions change several values at once

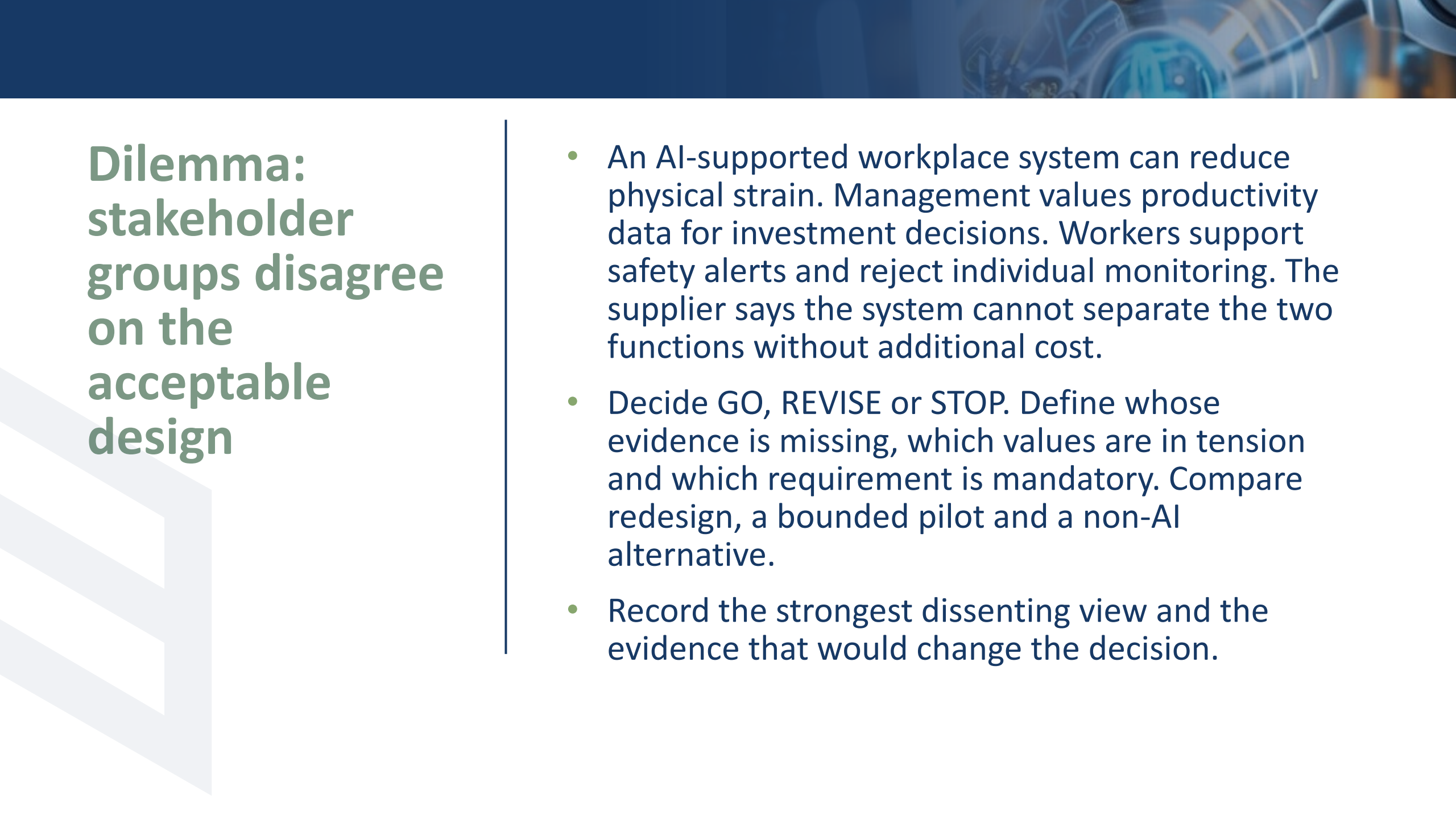
- SMOTE or class weighting may improve rare-class attention while affecting calibration. Transfer learning can reduce data and compute while importing unsuitable representations. Augmentation can improve robustness or create unrealistic cases. PCA can reduce complexity while hiding features relevant to a subgroup. LIME can aid local explanation while remaining approximate.
- Compare each intervention with a baseline. Test intended benefit, side effects, subgroup performance, uncertainty and operating limits.
- The tool is a design option, not evidence that the value has been achieved.

Applied task:

Conduct a structured values investigation

1. Choose hiring, historic photographs, data analytics or heart pumps.
2. Define the engineering decision and a non-AI alternative.
3. Map direct and indirect stakeholders, power and possible exclusion.
4. Define two relevant values in context and identify a tension.
5. Specify conceptual, empirical and technical evidence.
6. Create two feasible design options and mandatory thresholds.
7. Select an option, requirement, acceptance test, owner and review trigger.

Output: one page values investigation and a two-minute design defence.



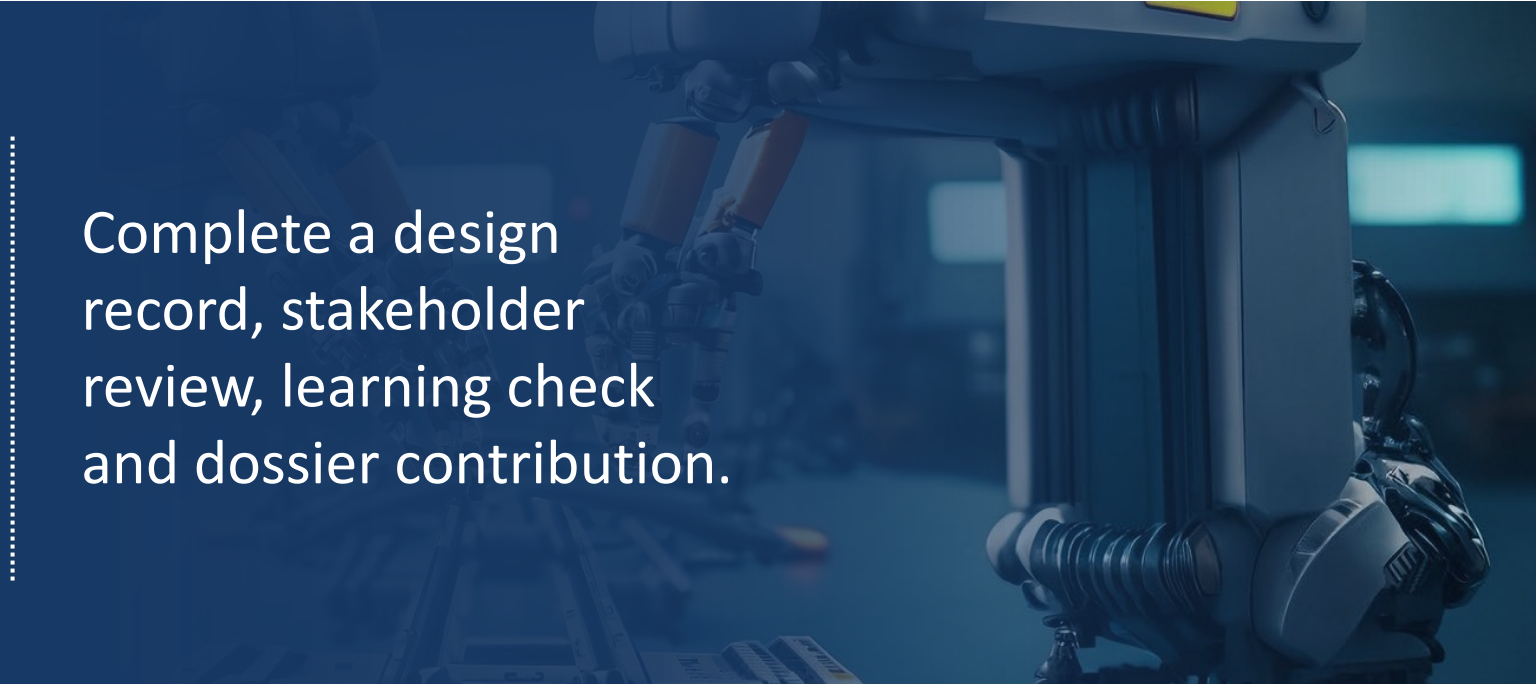
Dilemma: stakeholder groups disagree on the acceptable design

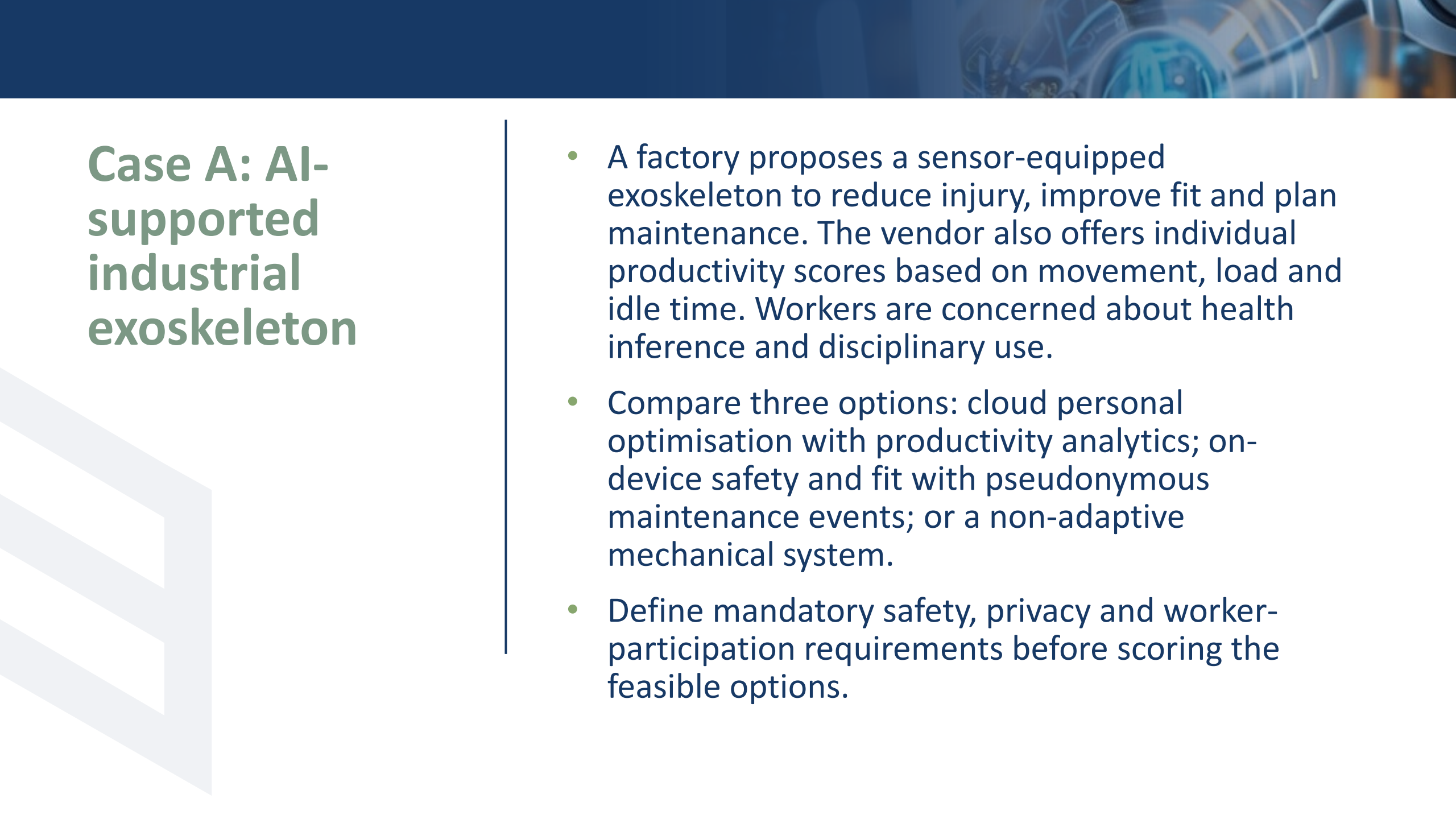
- An AI-supported workplace system can reduce physical strain. Management values productivity data for investment decisions. Workers support safety alerts and reject individual monitoring. The supplier says the system cannot separate the two functions without additional cost.
- Decide GO, REVISE or STOP. Define whose evidence is missing, which values are in tension and which requirement is mandatory. Compare redesign, a bounded pilot and a non-AI alternative.
- Record the strongest dissenting view and the evidence that would change the decision.

04

Defend a Values-Based Decision

Complete a design
record, stakeholder
review, learning check
and dossier contribution.





Case A: AI-supported industrial exoskeleton

- A factory proposes a sensor-equipped exoskeleton to reduce injury, improve fit and plan maintenance. The vendor also offers individual productivity scores based on movement, load and idle time. Workers are concerned about health inference and disciplinary use.
- Compare three options: cloud personal optimisation with productivity analytics; on-device safety and fit with pseudonymous maintenance events; or a non-adaptive mechanical system.
- Define mandatory safety, privacy and worker-participation requirements before scoring the feasible options.

Completed values-based design record: industrial exoskeleton

DECISION REVERSE: Select on-device safety and fit support; exclude individual productivity scoring.

STAKEHOLDER EVIDENCE: Workers, safety representatives, maintenance staff and supervisors prioritise fit, safe loading, clear alerts and freedom from disciplinary inference.

THRESHOLDS: Mechanical safety limits pass; no productivity or attendance use; worker representation approves the data boundary; local manual use remains available.

REQUIREMENTS AND TESTS: Process joint angles and load locally; delete raw motion buffers; pseudonymise maintenance events; test load limits, alert timing, fallback, usability and report contents.

REVIEW: Safety engineer owns load limits; data lead verifies prohibited use; worker representative reviews reports. Reopen after function, data, workplace or supplier change.



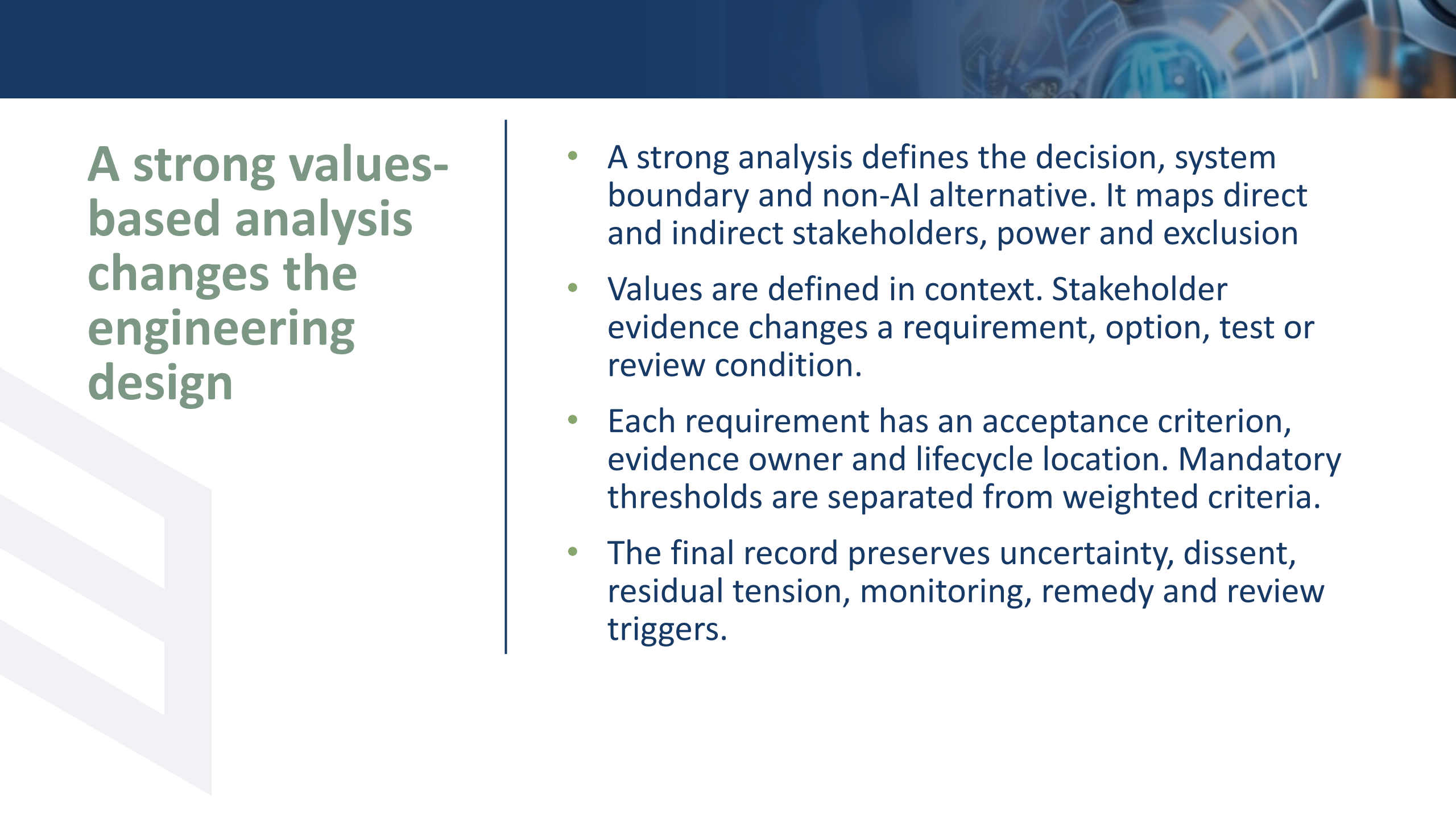
Case B: district-heating demand control

- A municipality plans AI control of district-heating supply using weather, building sensors and demand forecasts. The design can reduce fuel use, while conservative control may protect comfort for poorly insulated homes. Tenants cannot choose the building sensors and some lack a direct complaint route.
- Compare energy efficiency, thermal comfort, affordability, privacy, resilience and public accountability. Include a rule-based and operator-led alternative.
- Define minimum temperature and recovery thresholds, sensor purpose limits, operator override, public reporting and a remedy route before deployment.

Team Activity:

Values-based design decision lab

1. Choose the exoskeleton, district heating or a project case. Assign systems engineer, stakeholder lead, evidence lead, operations representative, critical reviewer and recorder.
2. Define the decision, boundary and non-AI alternative.
3. Complete conceptual, empirical and technical investigations.
4. Translate two values into requirements and acceptance tests.
5. Set mandatory thresholds, then compare two or three feasible options.
6. Decide GO, REVISE or STOP and record safeguards, owners, residual tensions and review conditions.
7. Present a two-minute defence and revise after stakeholder challenge.



A strong values-based analysis changes the engineering design

- A strong analysis defines the decision, system boundary and non-AI alternative. It maps direct and indirect stakeholders, power and exclusion
- Values are defined in context. Stakeholder evidence changes a requirement, option, test or review condition.
- Each requirement has an acceptance criterion, evidence owner and lifecycle location. Mandatory thresholds are separated from weighted criteria.
- The final record preserves uncertainty, dissent, residual tension, monitoring, remedy and review triggers.

Worksheet: Values-Based Design Record

System, intended use and non-AI alternative:

Direct and indirect stakeholders, power and exclusion:

Contextual value definitions and tension:

Conceptual evidence and assumptions:

Empirical evidence, participants and limits:

Technical options and value effects:

[Sources]

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

https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf. <https://standards.ieee.org/standard/7000-2021.html>

Continues On Next Slide



Mandatory thresholds:

Requirements, acceptance tests and owners:

Decision: GO REVISE STOP

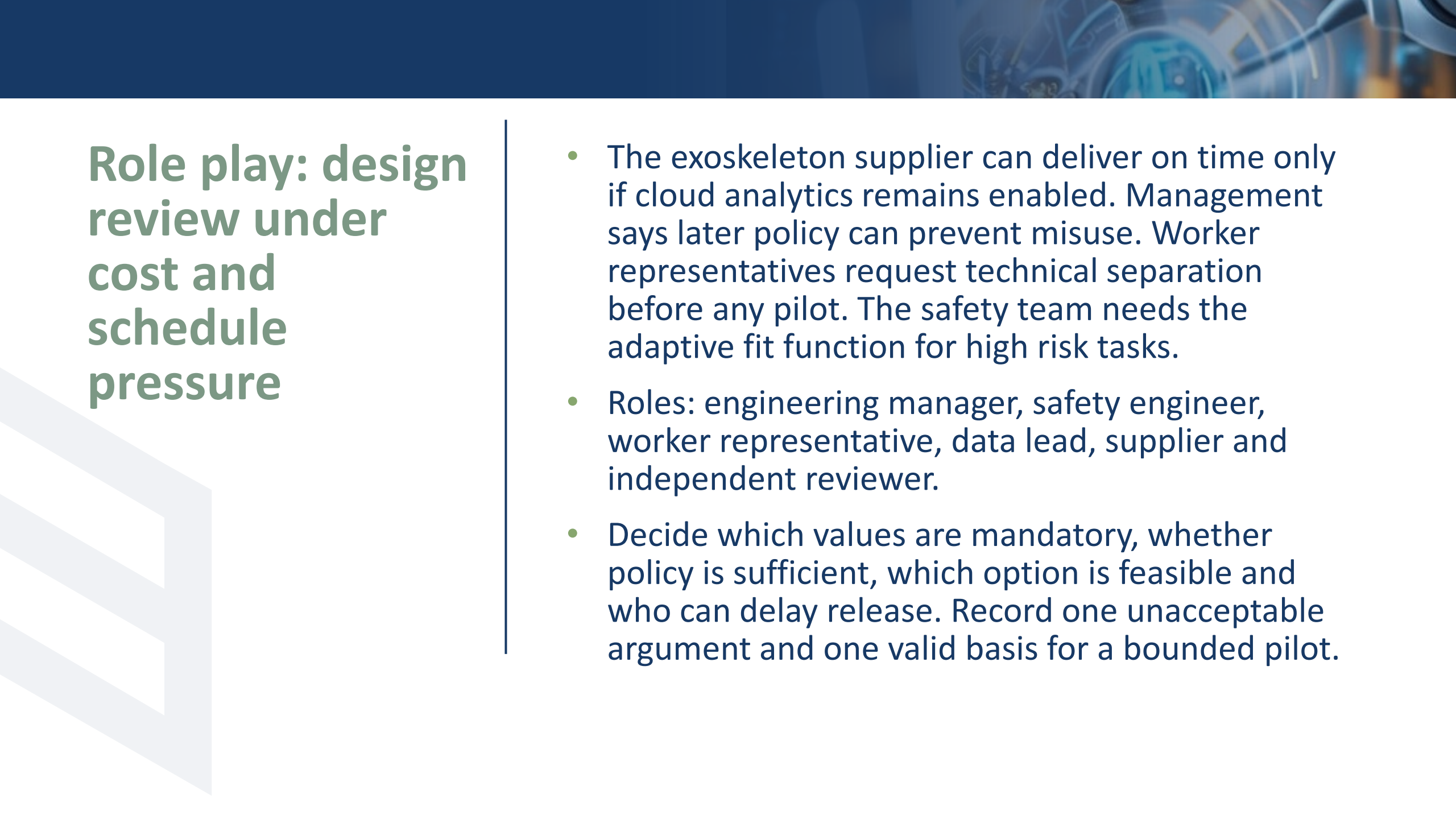
Safeguards, residual tensions and dissent:

Monitoring, remedy and review trigger:

[Sources]

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

https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf. <https://standards.ieee.org/standard/7000-2021.html>



Role play: design review under cost and schedule pressure

- The exoskeleton supplier can deliver on time only if cloud analytics remains enabled. Management says later policy can prevent misuse. Worker representatives request technical separation before any pilot. The safety team needs the adaptive fit function for high risk tasks.
- Roles: engineering manager, safety engineer, worker representative, data lead, supplier and independent reviewer.
- Decide which values are mandatory, whether policy is sufficient, which option is feasible and who can delay release. Record one unacceptable argument and one valid basis for a bounded pilot.

Individual reflection and transfer

- Write a short engineering argument.
- Which stakeholder or value was initially missing?
- Which evidence changed the problem definition or design?
- Which requirement made the value testable?
- Which threshold prevented an unacceptable trade-off?
- Who carries the remaining burden or risk?
- What change would reopen the decision?

Transfer the method to another engineering discipline and state one change in stakeholders, evidence or verification.

Quiz

1. How does a direct stakeholder differ from an indirect stakeholder?
2. What are the three Value Sensitive Design investigations?
3. Why should values be defined in context?
4. What connects a value with an engineering acceptance test?
5. Why are mandatory thresholds applied before weighted scoring?
6. Give one risk of a technical intervention such as SMOTE, PCA or post hoc explanation.
7. What evidence shows that stakeholder participation influenced the design?

[Sources]

- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>
- Original Module 5 deck supplied by the Ethical Engineer project consortium.

Quiz Answers & Teaching Rationale

1. Direct stakeholders interact with or receive decisions from the system; indirect stakeholders are affected through wider consequences.
2. Conceptual, empirical and technical investigations.
3. Safety, autonomy or fairness can mean different things across people, practices and systems.
4. A traceable value hierarchy: value, norm, requirement, acceptance criterion and evidence.
5. A weighted total must not cancel a serious legal, safety or rights failure.
6. Valid answers include calibration change, information loss, imported bias, instability or misleading explanation.
7. A recorded change to the problem, requirement, option, test or decision linked to stakeholder evidence.

Learning Outcome Check

Before finishing, confirm that you can:

1. Map direct and indirect stakeholders, power and exclusion.
2. Conduct conceptual, empirical and technical investigations.
3. Translate values into requirements and acceptance criteria.
4. Compare data, model, interface and system options using thresholds and evidence.
5. Produce a values-based design record with residual tensions and review.

Evidence:

Stakeholder map, value hierarchy, option comparison and design record.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>

Course Dossier Contribution

Add these Module 5 outputs to your ethical engineering dossier:

1. Stakeholder, power and exclusion map.
2. Contextual value definitions and value tension.
3. Conceptual, empirical and technical evidence.
4. Value hierarchy with requirements and acceptance tests.
5. Data, model and explanation intervention evidence.
6. Mandatory thresholds and option comparison.
7. Values-based design decision, safeguards and review trigger.

Connect this evidence with fairness, privacy and accountability work from Modules 2 to 4.

[Sources]

- Ethical Engineer approved KA220-HED application, WP3 Teaching and Learning.
- https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf
- <https://standards.ieee.org/standard/7000-2021.html>



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 5**



Co-funded by
the European Union