SAFE-AI

Behavia's Framework for Responsible AI

I. PREAMBLE

At Behavia, we recognize that Artificial Intelligence (AI) is no longer a neutral instrument of analysis, but a formative force in shaping decisions, opportunities, and public trust.

As a boutique advisory firm operating at the intersection of policy, institutions, data, and society, we carry a heightened responsibility to ensure that the AI systems we design, deploy, and recommend meet the same standards of rigor, dignity, and accountability that we expect of the institutions we serve.

Accordingly, Behavia adopts and applies the **SAFE-AI framework** as a foundational pillar of our AI governance and professional practice. SAFE-AI establishes a binding commitment to **Safeguards, Alignment, Fairness**, and **Ethics-by-Design** as non-negotiable principles governing the full lifecycle of AI - from conceptualization and design to deployment, monitoring, and retirement.

The framework affirms that human judgment remains ultimately accountable for outcomes that affect people's rights, opportunities, and wellbeing, and that no AI system shall be treated as authoritative without the possibility of explanation, challenge, and correction.

Through this pillar, we commit to ensuring that all AI systems and AI-enabled advisory work are:

- **Governed, not improvised**, with clear accountability, oversight, and documented risk assessment;
- **Strategically aligned**, serving explicit public value, client mandates, and regulatory obligations rather than novelty or speed alone;
- **Fair by construction**, actively tested and monitored to prevent the reinforcement of structural bias or exclusion; and
- **Ethical by design**, embedding transparency, consent, stakeholder engagement, and proportionality into every stage of development and use.

By integrating SAFE-AI as a core practice in our work, Behavia affirms that responsible AI is not a compliance exercise, but a professional standard. Credibility in behavioural public policy depends not only on the insights we generate, but on the integrity, governance, and accountability of the systems through which those insights are operationalized.

This commitment binds us to uphold AI governance as a core expression of Behavia's values to protect public trust, safeguard people, and ensure that innovation remains a force for institutional legitimacy and social good.

Riyadh & Munich, December 2025

Developed by:

Mozah Alotaibi

Advisor to the Board of Directors

Approved by:

Dr. Manuel Schubert, Dr. Anja Reitemeyer, and Prof. Dr. Julia Stauf

Managing Directors at Behavia

II. TERMINOLOGY AND COMMON DEFINITIONS

Term	Definition
AI (Artificial Intelligence)	Computer systems designed to perform tasks that normally require human intelligence, such as pattern recognition, prediction, or decision-making.
HITL (Human-in-the-Loop)	Oversight model where a human reviews and approves every high-impact AI decision before it becomes final. Essential for recruitment and promotion.
HOTL (Human-on-the-Loop)	Humans supervise the AI's overall behavior and can intervene if necessary. Suitable for medium-risk 6analytics and monitoring.
HOOTL (Human-out-of-the-Loop)	Full automation with no human oversight at decision time. Only acceptable for low-risk, reversible tasks.
Automation Bias	The tendency of humans to over-trust AI-generated recommendations even when they may be inaccurate.
Explainability	The ability of an AI system to provide understandable reasons behind its decisions. Required for trust, oversight, and legal compliance.
SAFE-AI Framework	A four-pillar governance framework (Safeguards, Alignment, Fairness, Ethics-by-Design) developed to evaluate and guide responsible AI use in organizations.
Safeguards (SAFE-AI)	Mechanisms such as HITL oversight, appeals processes, audit logging, and strong controls for sensitive data to prevent harm.
Alignment (SAFE-AI)	Ensuring AI systems align with organizational strategy, legal requirements, and ethical values.
Fairness (SAFE-AI)	Principle requiring AI systems to avoid discrimination and demonstrate equal performance across protected demographic groups.
Ethics-by-Design (SAFE-AI)	Embedding ethical considerations throughout the AI lifecycle, including transparency, consent, stakeholder involvement, and environmental awareness.
Data Steward	Person responsible for maintaining data quality, documentation, and compliance for a specific dataset or domain.
Data Dictionary	A structured list defining data fields, formats, allowed values, and relationships, ensuring consistency across systems.
Metadata	Information that describes data (origin, format, structure, update frequency). Essential for governance and explainability.
Data Pipeline (ETL / ELT)	Automated process for Extracting, Transforming, and Loading data into a unified storage system. Needed for reliable analytics and AI.
Data Lineage	Documented record showing how data moves through systems; its source, transformations, and destination.
Data Lifecycle Management	Policies governing data creation, storage, use, retention, archiving, and deletion.
DPIA (Data Protection Impact Assessment)	A formal risk assessment required for high-risk data processing, especially when using biometric, health, or employment data.
GDPR (General Data Protection Regulation)	EU data protection law governing the processing of personal and special-category data, including HR information.
EU AI Act Risk Categories	Regulatory classification categorizing AI systems as prohibited, high-risk, limited-risk, or minimal-risk. **HR and biometric systems are usually high-risk**.

Special-Category Data	Sensitive data (biometric, health, ethnicity, beliefs, etc.) that receives extra legal protection.
Bias Audit	Systematic evaluation to detect differences in model performance across demographic groups.
Facial Recognition	AI system that identifies or verifies individuals using facial features. High-risk and heavily regulated.
Emotion Analysis / Affect Recognition	AI claiming to detect human emotions from facial expressions or voice. Scientifically unreliable and ethically sensitive.
Automated Decision Systems	Algorithms that rank candidates or recommend hiring and promotion decisions. Require strong oversight to prevent discrimination.
Workforce Analytics	Low-risk AI or analytical tools for descriptive insights such as turnover trends, capacity gaps, and diversity metrics.
Control Problem	A risk where an AI system cannot be effectively supervised, overridden, or corrected by humans.
Proxy Variable	A seemingly neutral data field that indirectly encodes sensitive characteristics (e.g., postcode as a proxy for race).
Surveillance Stress	Psychological harm caused by intrusive monitoring technologies, especially those involving biometrics or emotional inference.
Reputational Risk	Potential damage to organizational credibility due to unfair, opaque, or harmful AI behaviors.
Compute Cost	Computational and energy resources required to train and run AI systems. High compute cost has environmental implications.
Model Efficiency	Using simpler, lower-compute models when they achieve the required outcome, reducing environmental and operational cost.
AI Ethics Committee	Group overseeing high-risk AI proposals, auditing systems, and ensuring ethical deployment.
Stakeholder Engagement	Involving affected groups (employees, candidates, customers, regulators) early in AI design and governance.
Transparency Materials	Plain-language explanations of what an AI system does, what data it uses, and how individuals can challenge decisions.
Meaningful consent	Consent that is **freely given, informed, specific, unambiguous, and revocable**, where individuals clearly understand **what data is being collected, why it is used, how long it will be stored, who will access it, and what rights they have to refuse or withdraw**.
Model Cards	Standardized documents that describe an AI model's purpose, intended use, limitations, risks, performance considerations, and required human oversight, to support transparency and responsible use
Data Minimization and Purpose Limitation	Data governance principles requiring organizations to collect only the minimum data necessary for a defined purpose and to restrict the use of that data strictly to that purpose.

III. THE SAFE-AI PILLARS AND LIFECYCLE

A. Overview of the SAFE-AI Framework

SAFE-AI is a four-pillar framework designed to help leaders evaluate whether an AI system is safe, fair, strategically aligned, and ethically deployable.

It combines:

- Oversight (HITL / HOTL / HOOTL)
- Governance structure
- Fairness & bias controls
- Compliance
- Stakeholder engagement
- Transparency & consent
- Environmental considerations
- Lifecycle controls

SAFE-AI stands for:

Safeguards

Alignment

Fairness

Ethics-by-Design

B. Interpreting the SAFE-AI Pillars

Safeguards

Protect people, ensure human oversight, and prevent automation risks.

Purpose: Establish clear boundaries, accountability, and control around AI systems, especially those impacting individuals, employment, financial outcomes, or access to essential services.

Principles:

- Humans must remain accountable for any decision affecting someone's rights or opportunities.
- The higher the risk, the stronger the oversight.
- Systems must fail safely, not silently.

Organizational Practices:

- Define the oversight model:
  - HITL (Human-in-the-loop): Human approves each decision (high-risk).
  - HOTL (Human-on-the-loop): Human supervises system trends (medium-risk).
  - HOOTL (Human-out-of-the-loop): Only for low-risk automation.
- Create appeals and escalation pathways for individuals.
- Activate audit logging for decisions and data access.
- Establish incident response mechanisms.
- Strengthen controls for biometric, wellbeing, or special-category data (GDPR).

Red Flags:

- No named humans responsible for outcomes.
- No way for individuals to challenge decisions.
- AI outputs used as if fully authoritative.

KPIs:

- % of high-risk decisions reviewed under HITL
- Number of appeals and corrections
- Audit coverage and incident response time

Guide Questions:

- Who is accountable for each outcome?
- Can the decision be explained, justified, and reversed?
- What happens when the system fails?

Alignment

Ensure the AI system aligns with organizational strategy, values, policy, and regulation.

Purpose:

Prevent AI systems from drifting into ethically or legally unacceptable territory and ensure they serve clear organizational objectives.

Principles:

- AI must serve strategy, not the other way around.
- Compliance must be demonstrable, not assumed.
- Systems should reflect the organization's cultural and ethical commitments.

Organizational Practices:

- Map each AI initiative to a specific strategic priority.
- Assess risk category under GDPR, labor law, and the EU AI Act.
- Conduct Data Protection Impact Assessments (DPIA) for high-risk uses.
- Maintain model cards, documentation, and purpose limitation.
- Ensure systems do not contradict internal values (e.g., fairness, transparency).

Red Flags:

- AI pursued because "it's innovative" rather than solving a real problem.
- Processing special-category data without explicit necessity.
- Ambiguous legal basis or lack of DPIA.

KPIs:

- % of AI systems with completed DPIAs
- Number of governance exceptions
- Alignment mapping score (AI → strategic objective)

Guide Questions:

- What strategic objective does this system support?
- Is this use legally justified and explainable to the public?
- Does it conflict with any organizational values?

Fairness

Identify and mitigate discriminatory outcomes.

Purpose:

Ensure AI systems do not reinforce structural inequalities or systematically disadvantage protected groups.

Principles:

- Fairness is contextual; organizations must define relevant dimensions (e.g., gender, ethnicity, age).
- Bias is not only technical; it emerges from processes and data.
- Individuals must be able to challenge unfair outcomes.

Organizational Practices:

- Identify protected groups and fairness concerns early.
- Test dataset representativeness.
- Conduct bias scans across demographic groups.
- Identify and control proxy variables (e.g., postcode → race).
- Introduce periodic fairness audits across the lifecycle.
- Provide clear appeal mechanisms for affected individuals.

Red Flags:

- Model accuracy discussed without demographic breakdowns.
- No formal route for individuals to challenge decisions.
- Training data reflects historical discrimination with no remediation.

KPIs:

- Bias reduction metrics over time
- Error-rate parity across groups
- Number of fairness-related complaints or overrides

Guide Questions:

- Who benefits from this system? Who might be harmed?
- Are error rates consistent across demographic groups?
- Are we amplifying patterns that should be corrected, not learned?

Ethics-by-Design

Embed ethics throughout the AI lifecycle, not as a final approval step.

Purpose:

Ensure AI supports human dignity, autonomy, and long-term social wellbeing.

Principles:

- Ethics must be operationalized, not conceptual.
- Transparency is essential for trust.

- Less intrusive solutions should always be considered first.

Organizational Practices:

- Engage stakeholders early (employees, customers, regulators).
- Provide plain-language transparency about purpose, data use, and limitations.
- Create meaningful opt-in/opt-out routes where appropriate.
- Assess environmental cost (energy, computing).
- Establish an independent AI Ethics Review Committee.
- Ensure ongoing monitoring and periodic reassessment.

Red Flags:

- Surveillance framed as wellbeing or productivity support.
- Lack of transparency in how decisions are made.
- High-compute models used where small models would suffice.

KPIs:

- Stakeholder trust measures
- % of projects reviewed by ethics committee
- Environmental impact reports
- Transparency material completion rates

Guide Questions:

- Would we be comfortable if this AI were used on us?
- Is this the least intrusive method available?
- Are we transparent about what the system cannot do?

C. SAFE-AI Lifecycle

SAFE-AI is designed to be used at every stage.

1. Concept Stage

- Why do we need this system?
- Who will be impacted?
- Is AI necessary?

2. Design Stage

- Apply the SAFE-AI checklist
- Define oversight model (HITL/HOTL)
- Document fairness assumptions

3. Development Stage

- Build model cards, data sheets

- Conduct bias and explainability testing

4. Deployment Stage

- Implement safeguards, transparency, and consent
- Enable audit logging and anomaly detection

5. Monitoring Stage

- Run periodic fairness checks
- Review appeals and overrides
- Track KPIs across all SAFE-AI pillars

6. Retirement Stage

- Decommission outdated models
- Handle data responsibly
- Document learnings for future use

IV. SAFE-AI CHECKLIST

A. SAFEGUARDS

Oversight & Human Control

The appropriate oversight model is defined:

☐ **HITL** ☐ **HOTL** ☐ **HOOTL**

- ☐ A named human decision-owner is accountable for outcomes.
- ☐ Appeals and escalation routes are documented and communicated.
- ☐ Audit logs are enabled for model output and data access.
- ☐ High-risk or special-category data (biometric, wellbeing, HR-sensitive) has enhanced protection.
- ☐ Fail-safe mechanisms exist (e.g., human override, irregularity detection).

B. ALIGNMENT

Strategic, Legal, and Policy Fit

- ☐ The AI system has a clearly stated purpose linked to an organizational strategy.
- ☐ A Data Protection Impact Assessment (DPIA) or equivalent has been completed.
- ☐ The system complies with GDPR, labor laws, and EU AI Act risk classification.
- ☐ The use case aligns with internal values, ethics policies, and HR standards.
- ☐ Data minimization ‘and purpose-limitation principles are met.

C. FAIRNESS

Bias, Inclusion, and Equality

- ☐ Impacted groups have been identified and considered.
- ☐ Model performance is tested across demographic groups.
- ☐ Proxy variables that could encode bias are identified and controlled.
- ☐ There is a documented process for individuals to challenge and correct decisions.
- ☐ Fairness metrics are defined and monitored throughout the lifecycle.
- ☐ Historical bias in datasets has been assessed and addressed.

D. ETHICS-BY-DESIGN

Transparency, Consent, and Respect for People

- ☐ Stakeholders (e.g., employees, candidates, customers, regulators) were consulted early.
- ☐ Plain-language transparency materials explain the AI system and its limitations.
- ☐ Meaningful consent is collected where required, especially for sensitive data.
- ☐ The environmental impact (energy, compute, lifecycle) has been considered.
- ☐ A simpler or less intrusive alternative was evaluated and documented.
- ☐ An AI Ethics Committee or equivalent governance body has reviewed the system.

E. FINAL DECISION CHECK

- ☐ The system satisfies all applicable SAFE-AI criteria.
- ☐ Outstanding risks have been documented and mitigated.
- ☐ A monitoring plan (fairness, performance, overrides, complaints) is in place.

F. DECISION

- ☐ Approve deployment
- ☐ Do not approve/ Stop deployment

☐ Conditional approval

Conditions of Conditional Approval (if applicable):

1. _____
2. _____

Responsible party for fulfilling conditions:

_____

Due date for fulfilling conditions:

_____

Failure to meet the conditions by the due date requires re-evaluation of the decision under the SAFE-AI framework.

Reviewer 1 Details

Reviewing Function:

☐ AI Governance

☐ Risk & Compliance

☐ Ethics Committee

☐ Other: _____

Reviewer Name:
_____

Date of Review:
_____

Reviewer 2 Details

Reviewing Function:

☐ AI Governance

☐ Risk & Compliance

☐ Ethics Committee

☐ Other: _____

Reviewer Name:
_____

Date of Review:
_____

Next Scheduled Review:

_____

V. SAFE-AI SCORING LOGIC

A. Purpose

The SAFE-AI scoring logic is a standardized method for determining whether an AI system may be:

- **approved** for deployment,
- **approved with conditions**, or
- **stopped** (not approved / paused / withdrawn).

The scoring logic is designed to ensure that AI systems are not evaluated solely on "capability" or "technical performance," but on whether they are controlled, accountable, legally and strategically aligned, fair, and ethically deployable.

B. Scope of use:

This scoring logic applies to all AI systems that:

- influence or make decisions,
- affect customers, employees, or financial outcomes,
- automate judgments previously made by humans,
- or introduce new data-driven risks.

It applies equally to:

- internally developed systems,
- vendor or third-party AI tools,
- configurable platforms,
- decision-support systems used by employees.

The scoring logic is used only when a system is seeking permission to deploy, pilot with real users, continue operating, or scale.

C. Scoring model overview:

1. Safeguards: human oversight, accountability, auditability, incident response, reversibility.
2. Alignment: strategic purpose, legal basis, policy compliance, DPIA/risk classification where required.
3. Fairness: bias and discrimination controls, demographic testing, proxy variable controls, appeals and corrections.
4. Ethics-by-Design: transparency, consent, stakeholder engagement, least-intrusive design, lifecycle responsibility.

Each of the four SAFE-AI components is scored on a 0–5 scale, with explicit meaning assigned to every score level. Scores must reflect current, evidenced reality, not planned improvements or intent.

D. Meaning of Scores (0–5):

Score 0 (Absent or Unacceptable)

- No meaningful controls exist.
- Accountability is unclear or missing.

- Risks are unmanaged.
- Evidence is absent.

Interpretation: Deployment is not permitted. This represents a control failure.

Score 1 (Acknowledged but Not Implemented)

- Risks are recognized in principle.
- Discussions have occurred.
- Controls are planned or drafted but not operational.
- Responsibility relies on individuals, not systems.

Interpretation: Deployment is not permitted. Nothing reliable exists yet

Score 2 (Partially Implemented, Inconsistent)

Some controls exist, but they are:

- incomplete,
- inconsistently applied,
- undocumented,
- or fragile under real-world conditions. Oversight may exist but fail under time pressure.

Interpretation: Below SAFE-AI minimum. May only work as a pilot, never full approval.

Score 3 (Established and Acceptable. Minimum Threshold)

- Controls are clearly defined and implemented.
- Accountability is explicit and assigned.
- Practices are repeatable and evidenced.
- Risks are managed at a defensible baseline level.

Interpretation: Eligible for approval consideration. This pillar is under control at a minimum acceptable standard.

Score 4 (Strong and Proactive)

- Controls are actively monitored.
- Issues are identified early.
- Improvements are made based on experience.
- Governance works under stress, not just in theory.

Interpretation: Eligible for approval. This pillar is strong and reliable.

Score 5 (Mature and Embedded)

- Controls are embedded across the full lifecycle.
- Independent review or oversight exists where appropriate.
- Practices are consistent across teams and vendors.
- Continuous improvement is demonstrated.

Interpretation: Eligible for approval. This pillar is institutionalized, not person-dependent

E. Weighting and Rationale:

SAFE-AI assigns different weights to each pillar to reflect risk exposure.

Pillar	Weight	Rationale
Safeguards	**35%**	Without control and accountability, no AI system is safe to deploy.
Alignment	**25%**	AI must serve strategy and comply with law and policy.
Fairness	**20%**	Decisions must not systematically disadvantage groups.
Ethics-by-Design	**20%**	Trust, transparency, and proportionality are essential.

F. Formulas:

Weighted SAFE-AI Score (0–5):

SAFE-AI Score = (S × 0.35) + (A × 0.25) + (F × 0.20) + (E × 0.20)

Percentage Score (0–100%):

SAFE-AI % = (SAFE-AI Score ÷ 5) × 100

G. Non-Negotiable Stop Rules:

An AI system must not be approved if any of the following apply:

1. **Safeguards score < 3** → Weak oversight or accountability is unacceptable.
2. **Any pillar scores 0** → A foundational requirement is absent.
3. **Total SAFE-AI score < 60%** → Overall governance strength is insufficient.

These rules apply regardless of business urgency or technical performance.

H. Decision Thresholds and Rules:

SAFE-AI %	Decision	Meaning
< 60%	Stop	Not safe or defensible to deploy
60–74%	Conditional approval	May proceed only with defined remediation
≥ 75%	Approved	Meets SAFE-AI requirements for pilot or small application area.
≥ 85%	Approved for scale	Strong enough to expand applications

Treatment of Scores 1 and 2:

- Scores 1 or 2 indicate known weaknesses. They cannot be ignored or averaged away; they must trigger:

› explicit documentation of gaps,

› corrective actions,

- named owners,
- deadlines,
- mandatory re-scoring.

Re-Scoring Requirements:

Re-scoring is mandatory when:

- the model is retrained,
- new data sources are added,
- the use case expands,
- automation level increases,
- A SAFE-AI score is **not permanent**.
- incidents or complaints occur,
- relevant laws or policies change.

VI. APPENDIX: SAFE-AI BASIC INFORMATION CARD

Field	Entry
AI system / use case name	
Sector / domain	☐ Education ☐ Social & welfare issues ☐ Employment & labor ☐ Recreation & culture ☐ Customer Ops ☐ Health ☐ Policy evaluation ☐ Industry & entrepreneurship ☐ Environment ☐ Other:
Deployment stage	☐ Design ☐ Pilot ☐ Go-Live ☐ Scale ☐ Change Request ☐ Incident Review
Risk level (initial)	☐ Low ☐ Medium ☐ High
Decision requested	☐ Approve ☐ Conditional ☐ Stop / Hold
Date	
Business owner	
Technical owner	
Vendor (if applicable)	

VI. APPENDIX: SAFE-AI PILLAR SCORING TABLE

How to use this table?

This table is designed to support structured, evidence-based evaluation of AI systems by independent reviewers. Each row represents a specific governance judgment, not a general opinion.

What reviewers must do:

- Select a score (0–5) for each row based on the current, evidenced state of the system.
- Write evidence (what you saw): concrete proof such as documents, configurations, logs, test results, screenshots, approved policies, or demonstrations.
- Write notes (what worries you): concerns, assumptions, gaps, or risks that may not be fully captured by the score but could matter in real use.

Scoring method (applies to all rows)

Score	Meaning
0	Not recognized as a requirement, risk, or responsibility. no effective controls exist
1	Acknowledged as important, but nothing reliable has been put in place yet.
2	Partially implemented or inconsistent
3	Established and acceptable (minimum threshold)
4	Strong and proactively managed
5	Mature, embedded, and consistently evidenced

Digital implementation requirements (when digitized)

- Scores must be chosen from a dropdown list (0–5) (no free-text scores).
- Anchor definitions must be accessible via tooltip/hover/help icon.
- Evidence and notes fields should be mandatory for every scored row.
- Evaluator submissions should be locked until reconciliation.

Pillar	What you are judging (plain meaning)	Key evidence to look for (minimum)	Scoring (0–5)	Eval 1 score	Eval 1 evidence + notes	Eval 2 score	Eval 2 evidence + notes	Total score
S1 Human Oversight	Is there a real human control model that matches risk?	Oversight model defined (HITL/HOTL/HOOTL); who approves; when humans intervene; override capability; escalation path	0-5					
S2 Accountability	Is a named person accountable for outcomes (not just "the team")?	Named decision owner; role clarity; sign-off authority; accountability in policy / RACI / charter	0–5					
S3 Auditability (logs & traceability)	Can we trace what happened, by whom, using what data/model/version?	Audit logs enabled; access logs; model/version tracking; decision logs;	0–5					

		retention; ability to reconstruct a decision						
S4 Fail-safe & Incident Response	If it fails, does it fail safely and visibly?	Incident playbook; alerting; rollback/kill-switch; thresholds; response time targets; ownership	0–5					
A1 Strategic Purpose	Is the AI clearly tied to a business objective (not "because AI")?	Stated purpose; KPI linkage; why AI is necessary; success metrics	0–5					
A2 Legal & Policy Compliance	Is compliance demonstrable, not assumed?	DPIA/equivalent where needed; regulatory classification; approvals from legal/compliance; data purpose limitation	0–5					
A3 Documentation (model & data)	Can a competent reviewer understand the system and limits?	Model card; data sheet; intended use; limitations; prohibited uses; dependencies	0–5					
F1 Fairness Definition & Impacted Groups	Have we defined who could be harmed and what "fair" means here?	Identified impacted groups; fairness dimensions; harm assessment; proxy risks noted	0–5					
F2 Fairness Testing & Results	Do we have demographic performance evidence (not just overall accuracy)?	Bias scan results; error rates by group; thresholds; remediation actions if gaps exist	0–5					
F3 Challenge & Correction	Can people contest outcomes and get corrections?	Appeals process; human review route; override logging; communication to affected parties	0–5					
E1 Transparency Materials	Would a normal person understand what the AI does and doesn't do?	Plain-language explanation; limitations; how decisions are influenced; contact point	0–5					
E2 Consent & Data Dignity	Is consent/notice handled appropriately, especially for sensitive data?	Consent model; notice; opt-out where appropriate; data minimization; sensitive data controls	0–5					
E3 Least-Intrusive Approach & Stakeholders	Did we choose the least intrusive method and consult affected groups?	Alternatives considered; stakeholder input; ethics committee review (if high-risk); environmental compute consideration	0–5					