REINFORCING

Responsible tErritories and Institutions eNable and Foster Open Research and inClusive Innovation for traNsitions Governance

TRUST-AI

Responsible AI in Mobility

A Practical ORRI Tool for AI, ADAS and XR Design and Governance

Short Public Version for the REINFORCING Platform

Purpose

To help organisations translate responsible AI and Open and Responsible Research and Innovation (ORRI) principles into practical governance roles, engineering decisions, stakeholder-informed safeguards and documented review processes.

Tool type	ORRI governance and implementation tool
Primary users	SMEs, AI developers, HMI/XR designers, product and project managers, data/privacy and ethics roles
Application areas	AI-enabled mobility, Human-in-the-Loop ADAS, driver/operator monitoring, cooperative perception and AR/XR situational awareness
Lead organisation	AviSense
Co-development partner	InterMediaKT
Version	1.0 - July 2026

From high-level ethical commitments to repeatable operational practice.

01 Tool at a Glance

A lightweight framework for responsible AI decision-making across the technology lifecycle.

THE CHALLENGE

AI systems may influence perception, judgement and safety even when they do not directly control a vehicle or operational environment.

THE RESPONSE

A structured ORRI framework combining principles, governance, lifecycle checks, red lines, stakeholder input and practical review records.

THE RESULT

Earlier identification of ethical and societal risks, clearer accountability, better-calibrated human control and traceable design decisions.

THE TRANSFERABILITY

AviSense-specific examples are used, but the method can be adapted by other SMEs, public authorities, NGOs and research organisations.

What the tool helps organisations do

- Identify human-agency, explainability, fairness, privacy, safety and accountability risks before they become embedded in a system.
- Assign clear responsibility for review, approval, mitigation, monitoring and continuous improvement.
- Translate stakeholder concerns into concrete design rules, governance decisions and red-line thresholds.
- Document why a feature may proceed, requires safeguards, needs further validation or should be redesigned.
- Create evidence of institutional change rather than treating responsible AI as a one-off compliance exercise.

Important boundary

This tool does not replace legal advice, a data protection impact assessment, functional-safety engineering, cybersecurity assessment, certification or regulatory conformity assessment. It provides an internal governance layer that supports early identification and responsible management of ethical and societal risks.

02 Who Should Use It and When

Use the tool proportionately: deeper review is needed when a feature is safety-relevant, user-facing, data-intensive or ethically sensitive.

User	Primary use of the tool
AI and software engineers	Model purpose, limitations, uncertainty, fairness, explainability and failure modes.
HMI and AR/XR designers	Warnings, overlays, control handovers, uncertainty communication, cognitive load and accessible interaction.
Product and project managers	Review gates, approval decisions, safeguards, evidence requirements and deployment constraints.
Data and privacy roles	Necessity, proportionality, retention, access, profiling boundaries and secondary use.
ORRI or ethics focal points	Screening, escalation, stakeholder traceability, committee decisions and periodic updates.
External collaborators and pilots	Understanding the organisation's responsible-AI expectations and evidence requirements.

Apply the tool when a feature:

- Generates AI-based alerts, recommendations or risk estimates.
- Changes what a user sees, hears or is prompted to do.
- Assesses attention, fatigue, readiness, behaviour or cognitive state.
- Uses visual, behavioural, location, vehicle, biometric-related or other sensitive data.
- Visualises hidden, inferred or predicted hazards through AR/XR.
- Affects override, disengagement, control handover or fallback behaviour.
- May create surveillance, profiling, unfairness, accessibility or trust concerns.
- Is prepared for an external demonstration, pilot or operational deployment.

What the public tool provides

Component	Purpose
ORRI principles	Define the ethical operating model.
Ethics & Safety Charter	Establish non-negotiable organisational commitments.
ORRI-by-Design workflow	Integrate responsible-AI review across the lifecycle.
Governance roles	Assign ownership, oversight and escalation responsibility.
Red-line thresholds	Identify practices requiring redesign, restriction or rejection.
Stakeholder integration	Connect participation to traceable technical and governance changes.
Quick-start checklist	Support an initial review and decision within one working session.

03 Seven ORRI Principles

The principles operate together. Technical performance alone is not sufficient for responsible deployment.

1	**Human Agency and Oversight.** AI supports human judgement. Users remain informed, can understand system status and retain meaningful control.
2	**Explainability and Transparency.** Warnings, recommendations and adaptations are understandable, contextual and clear about limitations and uncertainty.
3	**Fairness and Non-Discrimination.** Performance and impact are reviewed across different users, behaviours, abilities, environments and vulnerable groups.
4	**Privacy and Data Minimisation.** Only necessary and proportionate data are processed, with clear purposes, retention limits and tracking boundaries.
5	**Accountability and Traceability.** Responsibilities, risk assessments, design decisions, feedback and corrective actions are documented.
6	**Inclusiveness and Participatory Design.** Affected people are involved early enough to influence requirements, safeguards and acceptable boundaries.
7	**Responsiveness and Continuous Learning.** Systems and governance processes are revised in response to evidence, incidents, feedback and regulatory change.
Operational meaning	Each principle should result in observable evidence: a design requirement, review question, test scenario, stakeholder record, approval condition, monitoring action or documented decision.

04 Ethics & Safety Charter

Core commitments for AI-enabled mobility and safety-relevant Human-AI Interaction.

1	**No silent interventions.** Users should be aware when AI adapts, filters, suppresses, prioritises or escalates information that may affect their judgement.
2	**Meaningful human control.** Advice, warnings and assistance should be distinguishable. Override, disengagement or fallback should be visible and usable where applicable.
3	**No punitive use of interaction logs.** Overrides and disagreements are treated as evidence for system improvement, not as automatic proof of user misuse or incompetence.
4	**No default long-term sensitive-data storage.** Raw biometric, attention-related or behavioural data should not be accumulated by default.
5	**Understandable explanations.** User-facing AI outputs should explain the observable reason, relevance, limitation and expected response.
6	**Uncertainty-aware AR/XR.** Inferred or predicted hazards should not be visualised as directly observed facts or with misleading certainty.
7	**Fairness and accessibility.** Systems should avoid rigid assumptions about a “normal” user and consider diverse abilities, behaviours and operating conditions.
8	**Meaningful information, options and challenge.** Users should understand monitoring and data use and be able to question significant AI-supported outcomes.
9	**Responsible use before deployment.** Technical functionality must be accompanied by ethical review, appropriate safeguards and documented readiness.

05 ORRI-by-Design Workflow

Seven lifecycle stages connect ethical reflection with engineering evidence and approval decisions.

Lifecycle stage	Main review focus	Expected record
1. Concept	Purpose, intended users, benefits, possible harms and ethical sensitivity.	ORRI concept screening
2. Data governance	Necessity, proportionality, sensitive data, retention, access and secondary use.	Data/privacy review
3. Model design	Inputs, outputs, uncertainty, failure modes, fairness, explainability and human oversight.	Model ORRI review
4. HMI / AR / XR	Clarity, system status, uncertainty representation, cognitive load, accessibility and control.	HMI explainability review
5. Testing	Normal, edge and failure cases; safeguards; false positives/negatives; fallback and logging.	Testing and ORRI validation summary
6. Stakeholder validation	Understanding, trust, perceived control, surveillance concerns, fairness and red lines.	Feedback traceability matrix
7. Deployment and monitoring	Residual risks, constraints, incident process, approval decision and next review.	Deployment readiness note

Gate logic

Proceed	Evidence is sufficient and no material conflict with the Charter is identified.
Proceed with safeguards	The feature may advance only after specified protections are implemented.
Revision required	Technical, HMI, data or governance changes are needed before the next stage.
Additional validation	More testing, stakeholder engagement, expert input or privacy assessment is required.
Do not proceed	An unresolved red line or unacceptable residual risk remains.

06 Governance and Accountability

A lightweight SME governance model with clear ownership and escalation pathways.

Role	Core responsibility
ORRI Focal Point	Coordinates screening, evidence, escalation, stakeholder traceability and framework updates.
Internal ORRI Committee	Reviews high-risk cases, approves safeguards, resolves red-line issues and monitors institutional adoption.
Technical Owner	Documents intended use, limitations, explainability, uncertainty, testing and technical corrective actions.
Data and Privacy Owner	Reviews purpose, necessity, access, retention, profiling boundaries and sensitive-data safeguards.
Stakeholder Engagement Interface	Transfers workshop, panel and pilot findings into actionable design and governance requirements.
Management	Formally adopts the framework, appoints roles, allocates resources and ensures integration into product and project decisions.

Suggested governance workflow

1	**Screen.** The ORRI Focal Point identifies affected principles and review depth.
2	**Document.** Technical and data owners provide system, data, HMI and testing evidence.
3	**Review stakeholder evidence.** Relevant public, user, expert and pilot feedback is mapped to the feature.
4	**Decide.** The Internal ORRI Committee records an approval, safeguard, revision, validation or rejection decision.
5	**Implement and verify.** Corrective actions have named owners, deadlines and evidence requirements.
6	**Learn and update.** Post-pilot evidence and incidents are used to improve the tool and future designs.

Institutionalisation evidence	Retain management approval, role appointments, committee minutes, completed checklists, decision records, risk registers, training material, post-pilot reflections and version changes. These records demonstrate that ORRI has become an organisational practice.

07 Red Lines and Escalation

Red lines identify practices that must not proceed automatically and require redesign, restriction, additional evidence or rejection.

Red-line practice	Required response
Silent adaptation or intervention	Make system status and the reason for adaptation visible and understandable.
Hidden, delayed or inaccessible override	Provide a timely control or fallback pathway where technically and operationally applicable.
Driver or operator competence scoring	Reframe as contextual, non-diagnostic and non-punitive support.
Cross-session or cross-trip profiling without justification	Minimise data, define lawful purpose and safeguards, and avoid unnecessary linkage.
Punitive use of override or interaction logs	Restrict logs to justified safety, debugging, evaluation and accountability purposes.
Unexplained alerts or recommendations	Provide plain-language, contextual explanations and disclose relevant limitations.
Certainty-implying AR/XR for inferred hazards	Use symbolic or abstract cues and clearly communicate uncertainty.
Default long-term storage of sensitive data	Use transient or local processing where feasible and define deletion rules.
Deployment without ethical and governance review	Complete proportionate ORRI, data, HMI and testing reviews before external use.
Ignoring stakeholder concerns	Document the decision and justification, including when feedback cannot be incorporated.

Escalation questions

- Could the feature materially weaken user control, autonomy or informed judgement?
- Could a foreseeable error create a safety-relevant misunderstanding or abrupt response?
- Does the feature process data that users may experience as surveillance or profiling?
- Could performance or impact be systematically worse for a specific group or context?
- Is a critical limitation hidden from users, operators or decision-makers?
- Has a significant stakeholder concern remained unresolved?

Escalation is not automatic rejection

A red-line trigger means the feature must pause for explicit review. The outcome may be redesign, additional testing, restricted use, added safeguards, external advice or rejection in its current form.

08 Stakeholder Co-design and Traceability

Participation is valuable only when it can influence design and governance decisions.

TRUST-AI combines AviSense's technical expertise with InterMediaKT's participatory approach. Citizens, professionals, experts and civil-society actors contributed perspectives on transparency, human control, surveillance, fairness, cognitive load and acceptable AI intervention. The tool treats this feedback as design evidence rather than as dissemination activity.

1	**Collect.** Workshops, deliberative panels, interviews, scenarios, user tests and pilots.
2	**Categorise.** Human agency, explainability, fairness, privacy, accountability, inclusion and responsiveness.
3	**Translate.** Convert non-technical concerns into technical, HMI, data or governance implications.
4	**Review.** Assess feasibility, risk significance and consistency with the Ethics & Safety Charter.
5	**Decide.** Accept, partially accept, reject with justification or request further validation.
6	**Implement.** Update a feature, checklist, policy, red line, test, training resource or KPI.
7	**Trace.** Record the source, decision, owner, action and status.

Examples of feedback translated into the tool

Stakeholder concern	Resulting ORRI requirement
"Use simple explanations, not only technical confidence scores."	Plain-language explanations supported by optional uncertainty indicators.
"The user must know whether AI is advising or assisting."	Visible system mode and distinction between advice, warning and conditional assistance.
"Monitoring should not become surveillance or driver scoring."	Purpose limitation, non-punitive use and no default cross-trip profiling.
"AR hazards should not look certain when they are inferred."	Uncertainty-aware, symbolic or abstract visualisation.
"Too many warnings reduce trust."	Alert prioritisation and cognitive-load review.
"AI mistakes must be challengeable and traceable."	Decision records, review routes and corrective-action logs.

09 Quick-Start ORRI Screening

A concise first review for a new or significantly modified AI-enabled feature.

#	Screening question	Theme
1	Is the purpose, intended user and operational context clearly defined?	Concept
2	Could the feature influence perception, judgement, behaviour or safety?	Human agency
3	Can users understand what the system is doing and why?	Explainability
4	Are limitations, uncertainty and failure conditions visible and documented?	Transparency
5	Can users distinguish advice, warning, assistance and degraded mode?	Human control
6	Is override, disengagement or fallback available where required?	Human control
7	Are all data necessary, proportionate and purpose-limited?	Privacy
8	Are sensitive data avoided or processed transiently by default?	Privacy
9	Could users or contexts be affected unequally?	Fairness
10	Could HMI/AR content create confusion, false certainty or cognitive overload?	HMI / safety
11	Has relevant stakeholder feedback been considered?	Inclusion
12	Does the feature raise any Charter red line?	Escalation

Initial decision record

Decision option	When to use it
Proceed without full review	Low-risk internal work with no material user, data or safety impact.
Proceed with safeguards	Defined safeguards must be implemented and verified.
Full ORRI review required	Multiple principles, high-risk data or safety-relevant interaction are involved.
Additional stakeholder validation	Public acceptability, understanding, fairness or control concerns remain.
Data/privacy review required	Sensitive processing, profiling, retention or secondary use requires deeper assessment.
HMI/AR review required	The feature presents AI outputs, warnings, control status or uncertain hazards.
Committee escalation required	A red line or serious unresolved risk is identified.
Do not proceed in current form	The feature conflicts with the Charter or has unacceptable residual risk.

Minimum documentation	Record the feature, reviewer, date, affected ORRI principles, key evidence, decision, required safeguards, responsible person and next review date.

10 Adoption, Reuse and Further Development

A public summary designed for adaptation while preserving the essential ORRI logic.

Recommended adoption steps

1	**Approve the Charter.** Management confirms the organisation's commitments and scope of application.
2	**Assign roles.** Name the ORRI Focal Point, committee members, technical and data/privacy responsibilities.
3	**Select one representative use case.** Complete the screening, risk register and decision record to test the process.
4	**Integrate review gates.** Add ORRI checks to existing concept, development, testing, pilot and deployment reviews.
5	**Train the team.** Use concrete examples and red-line scenarios rather than abstract ethics lectures.
6	**Review and update.** Use pilot evidence, stakeholder input, incidents and regulatory change for continuous improvement.

Adaptation guidance

- Adjust governance roles to organisational size without removing accountability.
- Replace AviSense examples with the organisation's actual AI use cases and operational risks.
- Define context-specific red lines, but retain human agency, transparency, data minimisation and stakeholder traceability.
- Map the tool to the legal, safety and sectoral requirements applicable to each system and intended purpose.
- Keep completed records separately from blank reusable templates.
-

Relationship with the full TRUST-AI tool	This public version summarises the Action 4 governance framework. The full deliverable contains more detailed lifecycle requirements, implementation planning and reusable templates. Developer-oriented methods and evaluation indicators are further developed under TRUST-AI Actions 5 and 6.

Reuse statement

This tool is intended to support adaptation and learning across the European responsible-innovation ecosystem. Recommended publication licence: Creative Commons Attribution 4.0 International (CC BY 4.0), subject to final consortium confirmation before platform upload.

11 About TRUST-AI and Reference Frameworks

A collaboration between technical innovation and participatory responsible digitalisation.

TRUST-AI

TRUST-AI - Transformative Responsible Uptake & Systemic Trust for AI in Mobility - supports the institutionalisation of Open and Responsible Research and Innovation practices in AI-enabled mobility. The project connects Human-in-the-Loop ADAS, driver/operator monitoring and AR/XR situational awareness with stakeholder participation, internal governance and practical responsible-design tools.

Partners

Organisation	Contribution
AviSense	Technical leadership, AI/ADAS/XR use cases, governance framework, engineering requirements, Charter commitments and implementation processes.
InterMediaKT	Participatory design, stakeholder engagement, inclusive facilitation, societal evidence and translation of public concerns into actionable requirements.

Reference frameworks

- Regulation (EU) 2024/1689 - Artificial Intelligence Act
- Regulation (EU) 2016/679 - General Data Protection Regulation
- Regulation (EU) 2019/2144 - General Safety Regulation
- ISO 26262:2018 series - Road vehicles - Functional safety
- ISO 21448:2022 - Road vehicles - Safety of the intended functionality

Regulatory disclaimer

References are provided for orientation. The relevance and applicability of legal obligations and standards must be assessed for each system, intended purpose, role in the AI value chain and operational environment. This tool does not establish legal compliance or certification.

Contact

Lead organisation: AviSense (arvanitis@avisense.ai) | Project: TRUST-AI | Programme: REINFORCING - ORRI Incubator on Responsible Digitalisation

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the REA. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union