

OECD AI Transparency Report

Organization: MGOIT (ROU)

Reporting Period: Q2 2026

Published: October 2, 2026

Section 0 – Your role in the AI system lifecycle

Responding as (please select all that apply):

- Application developer (/provider)

If any of the fields above are selected, please indicate your organization size:

Micro enterprise: fewer than 10 employees

Section 1 – Risk Identification and Evaluation

1. How does your organization define and/or classify different types of risks related to AI?¹

- Maintain an internal taxonomy, classification system, and/or set of risk area definitions
- Define enhanced safety measures for specific dangerous capability thresholds, that, once reached, require safeguards (e.g. capability threshold for dangerous Chemical, Biological, Radiological and Nuclear (CBRN), cybersecurity attacks, dangerous advanced autonomy capabilities, or other capabilities).
- Align with specific external frameworks, standards, reports, and/or guidance documents such as taxonomies, classification systems, and/or sets of definitions for risk areas (e.g. the NIST AI Risk Management Framework, the ISO/IEC 42001, the OECD AI Principles, the EU General-Purpose AI Code of Practice, the International AI Safety Report). (please specify): MGOIT aligns its AI risk management practices with the following external frameworks: (1) ISO/IEC 27001:2022 — the organisation is certified, and its information security management system (risk assessment, data encryption, access control, incident management) forms the operational foundation for AI-related risk controls; (2) the OECD AI Principles and the Hiroshima Process International Code of Conduct for Organisations Developing Advanced AI Systems — MGOIT reported under the HAIP Reporting Framework 1.0 (published in the OECD 2025 report 'How are AI developers managing risks?') and uses the Code of Conduct's action areas as a structuring reference for its internal AI governance practices; (3) the EU AI Act risk-based classification system, used to assess and categorise client-facing AI applications during scoping and design.

How does your organization define thresholds at which severe risks, unless adequately mitigated, may be considered "unreasonable"? Please describe the framework(s) most relevant to your organization's role and context. If your organisation does not use formal thresholds, you may describe instead the policies, decision criteria, or governance processes you apply when assessing severe AI risks. You may also describe: • Specific examples of contexts in which your models or systems could, in your

assessment, pose unreasonable or intolerable risk; • How mitigations are graduated by capability or propensity level; • How your organisation considers the specificities of different AI technologies, such as generative AI and agentic AI, including any specific risk assessment or governance mechanisms applied to agentic AI capabilities.

As an application developer and deployer, MGOIT does not train frontier models and therefore does not operate formal capability-based risk thresholds. Instead, we apply decision criteria and governance processes at the project and system level to determine when an AI risk would be unreasonable:

(1) Use-case gating at scoping. Before development begins, each AI feature is assessed against the EU AI Act risk classification and our internal criteria. We consider risk unreasonable — and decline or redesign the use case — when an AI system would make fully automated decisions with significant impact on individuals (e.g. financial, employment, or safety-relevant outcomes) without meaningful human review, or when the client's data governance cannot support the required level of data protection.

(2) Human-in-the-loop as a graduated mitigation. Mitigations scale with impact: low-stakes outputs (e.g. document drafting, information retrieval) ship with user-facing review; higher-impact outputs (e.g. financial calculations, offers, compliance-relevant content) require explicit human approval before any downstream effect. Edge cases detected in testing are routed to mandatory human review paths.

(3) Pre-deployment testing in secure environments. AI features are tested in isolated environments with anonymised or synthetic data before production exposure, including adversarial and edge-case testing for bias and incorrect outputs.

(4) Agentic AI specifics. Where we build agentic capabilities (multi-step workflows, tool use, integrations), agents operate under scoped permissions: restricted tool access, no autonomous write-access to production systems of record without an approval gate, full action logging for auditability, and rollback paths. Autonomy level is treated as a risk dimension in itself — broader autonomy requires stricter approval gates and monitoring.

(5) Escalation and stop criteria. If testing or production monitoring reveals behaviour that cannot be mitigated to an acceptable level (persistent hallucination in a high-impact path, unmitigable bias, data leakage risk), the feature is withheld or rolled back until redesigned. This decision sits with the technical leadership (CTO and technical leads), not with delivery deadlines.

2. Both before and after deployment, what practices does your organization use to identify and evaluate vulnerabilities, incidents, emerging risks and misuse throughout the AI system lifecycle?

- Internal red teaming³
- Publicly available benchmarks and measurement tools
- Structured techniques to explore the behavior or performance of a model and/or application, such as model elicitation or bottleneck assessments
- External model, system, or application-independent research initiatives
- Internal model, system, or application-independent research initiatives
- Automated monitoring / remote status reporting (e.g., telemetry pipelines, automated alerts when risk thresholds are breached)

Please briefly describe your organization's decision-making process for deploying an AI model or system including: who has decision authority; what evidence and criteria are required (e.g.,

performance, safety/security, robustness, fairness, privacy, compliance, independent external testing, evaluation, validation and verification (TEVV) or assurance); contextual caveats or limitations required in interpreting risk evaluation metric results; what approvals/sign-offs are needed; and what safeguards (e.g., mitigations, monitoring, or phased release) must be in place to move into deployment.

Decision authority. Deployment decisions rest with the CTO, supported by the two technical leads (System Architect and Lead Software/AI Engineer). For client-facing systems, deployment additionally requires explicit client acceptance following user acceptance testing (UAT) — no AI feature moves to production on MGOIT's decision alone.

Evidence and criteria required before deployment:

- Functional and performance validation in a staging environment mirroring production, using anonymised or synthetic data;
- Edge-case and adversarial testing of AI outputs, including checks for bias, incorrect or fabricated outputs, and failure modes on malformed inputs;
- Security review consistent with our ISO/IEC 27001:2022 controls: access control, encryption in transit and at rest, secrets management, and data-flow review (what data reaches which model provider, under what terms);
- Privacy and compliance check: GDPR lawful basis, data minimisation, and anonymisation applied to any personal data in AI pipelines;
- Confirmation that graduated human-review mechanisms (per our risk criteria) are implemented and functioning, not merely designed.

Interpreting evaluation results. We treat evaluation metrics as context-bound: results obtained on test data are considered indicative, not guarantees, particularly for generative components where output quality varies with real-world input distribution. Known limitations (e.g. domains where the model underperforms, language coverage, edge cases routed to human review) are documented and communicated to the client as operating caveats, not hidden.

Sign-offs. Internal: technical lead responsible for the system signs off on testing evidence; CTO approves the release. External: client sign-off after UAT for client-operated systems.

Safeguards required at deployment:

- Phased release where feasible — limited user group or parallel run alongside the existing process before full cutover;
- Production monitoring and logging of AI outputs and agent actions, with alerting on anomalous behaviour;
- Defined rollback path allowing the AI feature to be disabled without taking down the surrounding system;
- Support and incident-response commitments (orientative SLAs of ~4h for critical and ~24h for major issues during the support period), under which AI-related incidents are triaged like any other production incident, with the added option of reverting to human-only workflow.

Following deployment, how do you use incident reports — including those shared by other organizations — to identify emerging issues, update risk assessments, and trigger corrective actions.

Post-deployment, AI-related incidents are handled through the same incident process as other production issues (ISO 27001 incident management, ~4h/~24h orientative SLAs during support periods). Each incident is logged, root-caused, and where relevant leads to corrective actions: prompt or model configuration changes, added human-review gates, or feature rollback. Recurring patterns

update our internal risk criteria and testing checklists, so lessons from one client project propagate to future scoping and QA.

We also monitor externally reported incidents — model provider advisories, security disclosures, and public AI incident repositories (e.g. the OECD AI Incidents Monitor) — and reassess our deployed systems when a reported failure mode is relevant to components or providers we use.

3. Does your organization have mechanisms to receive and/or share reports of risks, incidents or vulnerabilities by/with third parties?

Not applicable

4. If your organization has programs to incentivize risk(s) disclosure, incidents and systemic vulnerabilities, please share what type of programs you use?

- Recognition and reputation incentives
- Internal incentives and culture
- Operational incentives for customers and partners

Please provide public URLs and/or details on your incentive program(s). If none, write “Not applicable.”

No public program URL. As a small application developer, our incentive mechanisms are contractual and cultural rather than public programs: clients report issues (including AI-related ones) through a dedicated support channel with committed response SLAs (~4h critical / ~24h major), and internal reporting of defects and risks is standard engineering practice within our senior team, with no-blame handling.

5. Does your organization use or contribute to the development of international technical standards, best practices or tools for the identification, assessment, and evaluation of risks?

Contributes and uses

If contributes or uses, please select all that apply:

- ISO/IEC standards (e.g., ISO/IEC 27001, ISO/IEC 42001, ISO/IEC 23894, ISO/IEC 42005)
- NIST AI Risk Management Framework (AI RMF)
- OECD Due Diligence Guidance for Responsible AI

6. Does your organization collaborate with relevant stakeholders across sectors to assess and adopt risk mitigation measures to address risks that could have broad and potentially cascading socio-economic impacts? Please indicate which stakeholders you collaborate with, and whether collaboration is on a standing basis (Always-on), activated based on context (Ad hoc/as relevant), or there is no current collaboration.

	Always-on	Ad hoc / as relevant	No current collaboration
Relevant employees (security, legal/compliance, etc.)	☐		
Public authorities / regulators / AI safety institutes / other public sector bodies and inter-governmental coordination structures	☐		
Industry partners or industry fora		☐	
Users (end users, enterprise customers, developers using APIs)	☐		
Business customers / deploying organizations (if different from "users")	☐		
Supply-chain partners (cloud, data providers, etc.)	☐		
Academia / researchers		☐	
Independent auditors		☐	
Standards bodies / multi-stakeholder initiatives (OECD.AI; ISO/IEC; IEEE; AIID etc.)	☐		
Independent civil society organizations (e.g., for evaluations, audits, or policy review)		☐	
Media / general public (people or communities indirectly impacted)		☐	
Other(s) (please add any other relevant stakeholder groups):		☐	
Not applicable		☐	

Please describe examples of how your organization collaborates. Examples may include: participating in multi-stakeholder or industry initiatives; partnering with academia, civil society, or the public sector; establishing internal or joint governance structures with external experts or in partnership with other efforts to share knowledge and jointly address risks.

MGOIT participates in the OECD/G7 Hiroshima AI Process Reporting Framework since its first cycle (our HAIP 1.0 report was analysed in the OECD 2025 report 'How are AI developers managing risks?'), contributing an SME and application-developer perspective to an initiative otherwise dominated by large model developers. We continue under Framework 2.0.

At national and regional level, we share applied AI knowledge with industry: we run practical AI sessions for transport and logistics operators (private session with Romanian operators in June 2026, with a broader session planned for September targeting NL/DE/BE markets), covering both opportunities and risk/governance aspects of deploying AI in operations.

We also collaborate bilaterally with international partners on responsible AI integration in client projects, including sustainability and compliance-focused initiatives with partners in the DACH region.

Any further comments and implementation documentation.

No answer provided

Section 2 – Risk Management and Information Security

7. Across the AI system lifecycle, what steps does your organization take to manage risks and vulnerabilities? Please indicate which specific practices, tools or metrics your organization uses at each lifecycle stage. You may refer to tools or metrics present in the OECD.AI Catalogue of Tools and Metrics or suggest additions. Respondents may complete only the lifecycle stages relevant to their organisation's role. Stages that do not apply may be left blank.

Plan and design

Collect and process data

Build, integrate and interpret model

Verify and validate

Deploy

Operate and monitor

Please provide a consolidated description of your organisation's approach to risk management across the AI system lifecycle, including how the practices identified above operate together and any cross-cutting governance, review or escalation processes. Please also indicate how often these practices are reviewed and what types of events trigger reviews and updates. Where relevant, please distinguish between different types of AI systems (e.g. generative or agentic AI).

MGOIT's AI risk management runs through the full project lifecycle on a single governance chain: (1) scoping — use-case risk assessment (EU AI Act classification, internal criteria) with authority to decline or redesign unreasonable use cases; (2) design — graduated human-in-the-loop mechanisms and, for agentic systems, scoped permissions, approval gates on write actions, and full action logging; (3) pre-deployment — testing in secure staging environments with anonymised/synthetic data, edge-case and bias checks, security and GDPR review under our ISO/IEC 27001:2022 ISMS; (4) deployment — technical lead and CTO sign-off plus client UAT acceptance, phased release and rollback paths; (5) operations — monitoring, incident management with committed SLAs, and corrective actions feeding back into risk criteria and testing checklists.

Escalation runs to the CTO and two technical leads, who hold stop/rollback authority independent of delivery deadlines. Practices are reviewed at least annually alongside the ISO 27001 ISMS review cycle, and additionally triggered by: production incidents, relevant externally reported incidents or provider advisories, significant changes in model providers or capabilities, new regulatory requirements, and the start of each new AI project (scoping acts as a de facto review). Generative AI receives additional output-quality and hallucination controls; agentic AI receives stricter autonomy-based controls as described above.

Please describe any measures used to ensure data quality and fairness across the data collection and curation stage (e.g., dataset audits, filtering training data to remove harmful content or personal information, fine-tuning on curated data, use of moderation tools, assessment of limitations to model generalisability).

As an application developer, MGOIT does not train foundation models; our data work concerns curating client data for retrieval-augmented and fine-tuned components and application pipelines. Measures include: dataset review at ingestion (completeness, consistency, representativeness for the intended user population); removal or anonymisation of personal data before any data reaches AI pipelines, in line with GDPR minimisation; filtering of irrelevant or low-quality records that would degrade output quality; validation of embeddings/retrieval quality against known-answer test sets; and documented limitations where client datasets do not generalise (e.g. language coverage, historical bias in operational data), with those limitations communicated to the client and reflected in human-review gates.

8. What security measures are used in testing AI systems?

- Isolated and secure test environments
- Restrictions on outbound Internet or API access, including a list of permitted domains or APIs
- Isolation of staging environments (from production / operational environments)
- Human-in-the-loop controls for higher-risk test cases
- Access controls, including roles-based permissions
- Audit logging and monitoring
- Use of “AI-safe” test-data, as appropriate
- Use of synthetic, anonymized, or dummy data, as appropriate

8.A. Please indicate, if relevant, the team or role responsible for overseeing secure testing and whether this responsibility is shared across multiple teams and/or functions.

No answer provided

9. How does your organization protect intellectual property, including copyright-protected content?

9.A. Safeguards that promote lawful data use and transparency in AI training:

- Maintain a copyright policy that addresses data sourcing, content provenance, safeguards, and organizational roles and responsibilities
- Apply provenance controls to guard against training data being sourced from websites or repositories that have been determined to engage in large-scale piracy
- Refrain from the unlawful circumvention of technological protection measures in data sourcing
- Maintain policies or practices to respect copyright holders’ preferences through, where appropriate/applicable, mechanisms like Robots.txt “do not train” tags or other machine-readable protocols developed through voluntary, industry-led standards processes
- Engage in AI training that comprises a non-consumptive use of any underlying data
- Consult and/or retain an intellectual property attorney to protect, enforce, and comply with, intellectual property rights.

- Conduct data cleaning, qualification, and/or tokenization activities as part of a transformative data preparation process

9.B. Safeguards that reduce the risk of dissemination of outputs that may infringe copyright:

- Restricting outputs that raise ethical, human rights, or legal concerns—such as non-consensual intimate imagery, unauthorized deepfakes, or outputs that are substantially similar to copyrighted works.
- Filtering results using keyword or image-based restrictions to reduce risks of misuse
- Designating clear points of contact so creators or rights holders can raise concerns about infringing outputs

10. How do you protect privacy and prevent your systems from disclosing confidential or sensitive data across the AI lifecycle?

- Make information about data privacy, practices, and controls accessible to a diverse set of stakeholders
- Use security controls, including encryption at rest and in transit
- Provide organizational customers with controls to configure and implement privacy choices (e.g., data retention, logging, access, or processing settings)

Describe key privacy measures and data-protection standards applied.

- *Data minimisation and anonymisation: personal and sensitive data is anonymised or pseudonymised before entering AI pipelines; testing uses anonymised or synthetic data only.*
- *Encryption in transit and at rest for all data stores and AI-related data flows.*
- *Access control: role-based access, least-privilege, and secrets management for model provider credentials.*
- *Provider data-flow governance: for each project we document what data reaches which model provider under what contractual terms, and select providers/configurations with no training on customer data; where client requirements demand it, we deploy on-premise or private-cloud inference.*
- *Disclosure prevention at the application layer: retrieval components are scoped per-user/per-role so the system cannot surface documents the requesting user is not authorised to see; agent tool access is similarly permission-scoped; outputs in sensitive contexts pass human review before external use.*
- *Logging and audit trails of AI interactions, retained under defined retention policies.*
- *Contractual safeguards: NDAs and data-processing agreements with clients define confidentiality obligations covering AI components.*

11. How does your organization assess cybersecurity risks and implement policies to enhance the cybersecurity of advanced AI systems across the AI system lifecycle? Respondents may complete the checkbox subsections in 11.A–11.D or describe their overall approach under 11.E.

11.A. Governance, policy and assessments

- Maintain or update security policies aligned with recognized standards (e.g., NIST CSF, ISO/IEC 27001, SOC 2, CIS Critical Security Controls), and review them regularly in response to emerging risks or regulatory guidance

- Maintain vulnerability management processes (scanning, channels for external reporting, patching, remediation)
- Perform threat modeling that includes AI-specific risks (e.g., data poisoning, prompt injection)
- Integrate cybersecurity AI risks into broader cybersecurity risk management activities

11.B. Technical and operational controls

- Apply multi-factor authentication (MFA) and role-based access controls for advanced AI systems
- Encrypt AI assets (datasets, models, configurations) in storage and in transit as appropriate
- Monitor system activity and log access to detect anomalies or misuse

11.C. Protection of sensitive or high value AI assets

- Limit access to proprietary AI assets based on role or need-to-know
- Store model weights, datasets, and algorithms in secure, access-controlled environments
- Use hardware security modules (HSMs) or trusted execution environments (TEEs) where feasible
- Require NDAs or confidentiality agreements for staff and contractors with access

11.D. Insider risk protection detections

- Conduct pre-employment screening or role-sensitive vetting
- Monitor for unauthorized internal access or suspicious activity
- Provide staff training on cybersecurity best practices and AI-specific threats
- Implement insider-risk escalation or response procedures
- These responsibilities are delegated contractually to model or infrastructure providers

11.E. Please describe the main security measures your organization applies and indicate which are specific to particularly sensitive AI assets, where relevant.

Our security baseline is the ISO/IEC 27001:2022 ISMS, certified and externally audited, applied to all systems including AI components:

- *Risk assessment: periodic ISMS risk assessments, extended at project scoping to AI-specific threats (prompt injection, data leakage through model outputs, over-privileged agent access).*
- *Infrastructure and access: role-based access control, least privilege, MFA, encryption in transit and at rest, network segregation between environments (dev/staging/production), and hardened cloud or on-premise deployment per client requirements.*
- *Secure development: code review, dependency and vulnerability scanning, secrets management (no credentials in code), and isolated testing environments.*
- *Incident management: defined incident-response process with committed SLAs (~4h critical / ~24h major during support periods), logging and audit trails.*

Measures specific to sensitive AI assets:

- *Model provider API keys and inference credentials treated as high-sensitivity secrets (restricted access, rotation).*
- *Input/output controls against prompt injection and data exfiltration: input sanitisation on untrusted content, per-user scoping of retrieval so outputs cannot leak unauthorised documents.*
- *Agentic systems: scoped tool permissions, approval gates on write actions to systems of record, and complete action logging for auditability.*

- Client datasets and embeddings stores treated as confidential assets under the same encryption and access-control regime as source data.

12. What is your organization's vulnerability management process? Does your organization take actions to address identified risks and vulnerabilities, including in collaboration with other stakeholders?

- Internal vulnerability disclosure receipt and tracking processes
- Investigation of reported or identified risks and vulnerabilities
- Development of patches or other mitigations for confirmed risks or vulnerabilities
- Vulnerability disclosure program that is aligned with internationally recognized standards, e.g. ISO/IEC 30111 and 29147.
- Reporting to regulators / authorities when relevant
- Testing of patches or other mitigations

13. What post-deployment practices does your organization use to monitor and respond to vulnerabilities, incidents, emerging risks and misuse of advanced AI systems?

- Conduct post-deployment monitoring
- Perform root-cause analysis and apply lessons learned to improve future development and/or deployment practices
- Monitoring and emergency controls specific to agentic AI systems (e.g., ability to interrupt autonomous action sequences, override tool-use permissions, or revoke delegated authority)

14. Do you develop or deploy agentic AI systems (systems that can autonomously plan/act or use tools)?⁴

Yes

14.A. If yes, please indicate which categories of controls your organization applies. Please select all.

- Constraints on the agent's action space (e.g., tool-use permissions, environment sandboxing, allowlists/blocklists of resources, scope limits per task)
- Human oversight or interruptibility mechanisms (approval requirements for sensitive actions, escalation paths, interruption interfaces)

Please provide any further comments and implementation documentation.

No answer provided

Section 3 – Transparency Reporting on Advanced AI Systems

15. Does your organization publish public reports and/or technical documentation related to the capabilities, limitations, and domains of appropriate and inappropriate use of advanced AI systems? Are these clear and understandable by non-technical audiences?

Yes

15.A. If yes, which of the following information is included in your organization's publicly available documentation for any of your AI products?

- Risk management policies
- Summary of results from evaluations of safety, security, and societal risks including potential adverse impacts on individuals or groups (e.g. impacts on human rights, privacy or differential outcomes across user groups).
- Summary of results from red-teaming and other testing (e.g., adversarial evaluations, stress tests, and other pre-deployment safety assessments).

Please provide links or references showing where this information is made public.

No answer provided

15.B. If yes, how often are such reports usually updated?

Other (e.g., triggered by material changes, incidents, or new risks)

16. How does your organization share information with a diverse set of stakeholders (other organizations, governments, civil society and academia, etc.) regarding the outcome of evaluations of risks and impacts related to an advanced AI system?

- Public documentation
- Direct engagement with AI safety institutes, civil society or academic partners
- Regulatory reporting as relevant

Describe how evaluation results are shared and with which stakeholders (e.g. other organizations, governments, civil society and academia, etc). You may also indicate whether non-technical or plain-language summaries are provided for broader audiences.

As an application developer working on client-specific systems, our evaluation results primarily concern client deployments and are shared as follows:

- *Clients: full evaluation outcomes are shared directly — testing results, known limitations, operating caveats, and residual risks are documented and communicated in plain language during UAT and handover, so non-technical client stakeholders understand what the system can and cannot reliably do.*
- *Public/multi-stakeholder: we share our risk management practices and lessons learned through the OECD/G7 HAIP Reporting Framework (this report and our HAIP 1.0 submission), which is publicly available.*
- *Industry: we present applied lessons — including limitations and failure modes, not only successes — in practical AI sessions with industry operators (e.g. transport and logistics sector sessions).*

17. Does your organization publish public privacy policies addressing the use of personal data, user prompts, and/or the outputs of advanced AI systems?

No

18. How does your organization provide information about the sources of data used for the training of advanced AI systems, as appropriate, including information related to the sourcing of data annotation and enrichment?

- Provide information or tools to downstream actors on data quality
- Protected information sharing
- Regulatory reporting as relevant

Please provide links to any publicly available information.

I cant

19. Does your organization demonstrate transparency related to advanced AI systems through any other methods?

No

20. Does your organization publish statistical insights derived from anonymised, aggregated usage data to advance understanding of AI adoption and its economic impacts in areas including occupations and tasks by region, sector and country? To this end, do you contribute to international cooperation efforts?

Yes

If yes, please provide links or describe the scope of insights published, including where relevant contributions to understanding AI adoption in specific sectors (e.g. public administration, education systems).

I cant share this

Please estimate the total staff time and annual expenses associated with conducting compliance activities and include any further comments and implementation documentation (optional).

No answer provided

Section 4 – Organizational Governance, Incident Management, and Transparency

21. How has AI risk management been embedded in your organization's governance framework?

- Executive oversight
- Board of Directors oversight

- Senior management oversight

Please describe how AI risk management is integrated into governance structures, including: which roles or bodies have advisory versus decision-making authority; which roles or bodies have final accountability for AI risk-related decisions; and how often policies are reviewed or updated (e.g., following major incidents, regulatory changes, or material updates to systems).

AI risk management is embedded through executive and senior technical oversight, proportionate to our size (senior team of four). The CTO holds executive oversight and final accountability for AI risk decisions, including deployment approval and stop/rollback authority. The two technical leads (System Architect and Lead Software/AI Engineer) provide senior management oversight: they assess AI risks at scoping, design, and pre-deployment, sign off on testing evidence, and escalate concerns directly to the CTO. AI risk criteria are integrated into our ISO/IEC 27001:2022 ISMS, which is externally audited and reviewed at least annually, with additional updates triggered by incidents, regulatory changes, or material changes to systems or model providers.

22. Are relevant staff trained on your organization's governance policies and risk management practices?

Yes

If yes, describe training methods (e.g., onboarding modules, regular sessions, and confidential channels to report concerns).

Given our size, training is continuous and practice-based rather than formal modules: AI risk criteria, security practices, and lessons from incidents are covered in regular team sessions and project retrospectives, and applied directly through code review and scoping discussions where all four team members participate. New collaborators are onboarded on our ISMS policies (ISO/IEC 27001 requirement) including AI-specific data handling rules. Concerns can be raised directly to the CTO — in a four-person senior team, the reporting channel is immediate and no-blame by practice

23. Are steps taken to address reported incidents documented and maintained internally?

Not applicable

24. Does your organization share research and best practices on addressing or managing risk?

Yes

If yes, describe how research or good practices are shared (e.g., publications, conferences, collaborations with other sectors).

We share applied good practices rather than academic research: (1) through the OECD/G7 HAIP Reporting Framework — our public reports document our risk management practices as an SME application developer (HAIP 1.0, analysed in the OECD 2025 report, and this 2.0 report); (2) through practical industry sessions — hands-on AI sessions for transport and logistics operators covering deployment practices, risks, and governance (Romanian operators session in June 2026, broader

NL/DE/BE session planned for September); (3) through public professional content sharing lessons from production AI deployments, including what fails and why, not only successes.

25. Does your organization use international technical standards or best practices for AI risk management and governance policies?

Yes

If yes, please indicate which standards or frameworks apply:

- OECD AI Principles
- Industry-led best practices or codes of conduct
- ISO/IEC standards (e.g., ISO/IEC 27001, ISO/IEC 42001)

If yes, describe which international technical standards or best practices your organization uses.

ISO/IEC Standards 27001, OECD AI PRINCIPLELS, codes of conduct

Any further comments and implementation documentation.

No answer provided

Section 5 – Content Authentication & Provenance Mechanisms

26. What mechanisms, if any, does your organization put in place to allow end-users, where possible and appropriate, to know when they are interacting with an advanced AI system developed by your organization?

- User disclosure (e.g., text notice, verbal statement, or system label)
- Persistent visual or audio indicators (e.g., watermark, icon, chime, banner) during AI interaction
- Distinct branding or labeling for AI-generated content (e.g., “AI-generated” tag, metadata label)

Provide any additional details or context on how or when users are informed they are interacting with advanced AI systems (e.g. if such mechanisms are planned, in testing, or applied only to certain products).

No answer provided

27. Does your organization use content provenance detection, labeling or watermarking mechanisms that enable users to identify content generated by advanced AI systems?

No

Any further comments and implementation documentation.

No answer provided

Section 6 – Research & Investment to Advance AI Safety & Mitigate Societal Risks

28. Does your organization advance research and investment in the following areas?

- Security
- Safety
- Data quality and representativeness
- Explainability or interpretability
- System performance and outcomes (e.g. accuracy, consistency)
- Transparency
- Robustness or reliability
- Information integrity

If yes, please describe key research or investment initiatives in the areas selected above (e.g., dedicated safety teams, funding of research programs, contributions to open-source safety tools, publishing research, partnerships and collaborative research initiatives). You may also indicate whether related outputs are publicly shared, shared with partners or within consortia or not publicly disclosed.

As an SME application developer, our 'research and investment' takes the form of applied engineering investment rather than formal research programs. Across the areas selected: security and safety practices are developed and refined through our ISO/IEC 27001-certified ISMS and per-project AI threat assessment (prompt injection, data leakage, agent permissions); data quality, system performance, robustness, and explainability are ongoing engineering investment areas — we build and iteratively improve internal testing methods (known-answer retrieval validation, edge-case and bias testing, output-quality evaluation for generative components) reused across projects; information integrity and transparency are addressed through human-review gates, action logging, and documented system limitations communicated to clients.

Outputs are shared as follows: practices and lessons are public through our HAIP reports and industry sessions; internal testing methods and tooling are shared with clients and partners within engagements; client-specific results are not publicly disclosed (confidentiality).

29. Does your organization pursue research or investment related to socio-economic or environmental considerations in the development and deployment of AI?

- Environmental considerations (e.g., energy use, computational efficiency, resource use)

If yes, please provide examples of relevant research or investment initiatives for the areas selected above.

<https://www.libertatea.ro/stiri/green-deal-sustenabilitate-profit-companii-romania-timisoara-5774071> AI on ESG stuff

Any further comments and implementation documentation.

No answer provided

Section 7 – Advancing Human and Global Interests

30. Does your organization support any digital literacy, education or training initiatives to improve user awareness and/or help people understand the nature, capabilities and limitations of advanced AI systems?

Yes

If yes, describe any digital literacy or training initiatives your organization supports (e.g., user education, public AI-education programs, school or university partnerships, community workshops). You may also indicate whether these efforts include or prioritise underserved or underrepresented communities.

Our digital literacy efforts focus on practical AI education for industry professionals and SMEs: (1) hands-on AI sessions for transport and logistics operators — practitioner-level education on what AI can and cannot reliably do in operations, including risk awareness (June 2026 session with Romanian operators; broader session planned for September)

31. Does your organization pursue AI projects that aim to deliver societal or public benefits, including through partnerships with civil society and community groups?

Yes

If yes, please provide examples, including any collaborations with civil society, community groups, or efforts to address broader societal challenges.

No answer provided

Any further comments and implementation documentation.

No answer provided