

OECD AI Transparency Report

Organization: Malt Community SA (FRA)

Reporting Period: Q3 2026

Published: October 2, 2026

Section 0 – Your role in the AI system lifecycle

Responding as (please select all that apply):

- Model developer (/provider)
- Deployer
- Application developer (/provider)

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

Large enterprise: 250 or more employees

Section 1 – Risk Identification and Evaluation

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

- Consult with external stakeholders (e.g., from academia, industry groups, civil society, or standard development organisations) when developing or updating risk taxonomies
- 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): Malt risk management system primarily maintains an internal taxonomy to match Malt's AI system properties, guided by the EU AI Act standards. In addition, Malt's risk management system takes into consideration the NIST AI risk management framework, in particular the lifecycle stage and the characteristics of trustworthy AI.
- Maintain an internal taxonomy, classification system, and/or set of risk area definitions

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.

Malt's technology relies on an internal proprietary matching algorithm built on internal datasets and is used by over 100,000 companies to find talents from Malt's community of more than 1,000,000 qualified independent experts. Malt develops a proprietary recommendation system built on a small language model. In addition, Malt integrates third-party generative AI in its software to retrieve a set of relevant criteria from client's hiring needs (prompt – a pdf, a copy/paste of a job description, a long paragraph or just a few keywords), and agentic AI to assist the client throughout their search. All AI extracted hiring criteria or proposed actions require explicit client validation.

Malt maintains an internal AI risks register which includes risks whether they are algorithmic -derived from the AI model-, operational -derived from the use of the AI model-, or fundamental -impacts individual rights. The risk registers uses a traditional risk-mapping methodology reviewing the gravity scale, and the likelihood scale; and it integrates the lifecycle stages and vets the risks against the characteristics of trustworthy AI. It is maintained by the AI deployment committee. Risks related to historical data bias crystallization whereby the system replicates legacy (human) biases against certain demographic groups are considered serious/likely and are therefore unreasonable. Risks are analyzed, mitigated operationally and technically, and constantly monitored; they may prevent the model from being released/used.

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
- Independent third-party red teaming
- Adversarial testing to detect technical vulnerabilities
- Publicly available benchmarks and measurement tools
- Internal benchmarks, standard metrics (e.g., frequency, severity scores)
- Qualitative assessments (e.g., expert judgment, narrative scoring)
- Specific risk identification practices for vulnerable groups
- Impact assessments such as personal data protection assessment and fundamental rights impact assessments
- Internal model, system, or application-independent research initiatives
- External model, system, or application-independent research initiatives
- Incident reports, including reports shared by other organizations
- Vulnerability reports

Are any national AI safety / security institutes of the AISI network involved?

No

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.

The AI Deployment committee has decision authority when it comes to deploying AI models.

The committee oversees the lifecycle of in-house AI developments and external AI tools being built into Malt's products across two key pillars: (1) new model/feature: system risk classification, risks associated to the system, data protection and fundamental rights impact assessments, technical documentation check, transparency requirements check, and third-party vendor documentation (if applicable); (2) model/feature monitoring: review risks associated to the system, review incidents (if any), performance reviews and log retention audits (minimum 6-month retention).

Before submission to the AI Deployment committee, the machine learning/product owner must compile a standardized AI System dossier containing the evaluation framework which includes notably a list of criteria, a description of what is evaluated and the assessment method, the design specification, data protection and fundamental right impact assessments, and a information security report from Malt's information security team.

The committee holds binding authority on system deployment under three possible outcomes: (1) approved: sign-off for deployment under specified operational parameters; (2) conditional approval: requires specific remedies (e.g., additional human-in-the-loop controls, refined dataset balancing) prior to go-live; (3) rejected / suspended: immediate block on deployment or mandate to halt an active model due to unmitigated high risk or serious incident reporting.

Safeguards include a mandatory progressive deployment scaling from 5% to 25%, 50%, and 100% user traffic, with reviews at each increment; real-time drift and error monitoring triggers an automated switch to a deterministic, rule-based fallback system.

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.

Internal incident generates a detailed live incident document that records the incident description, chronological timeline, containment actions, and initial technical resolutions. Upon incident resolution, a post-incident analysis and retrospective are conducted. The incident response team investigates underlying root causes, identifies procedural or technical gaps, and formulates action plans.

Malt's AI deployment committee and information security team review internal incident reports related to the AI system, and monitor external incident reports in order to review the risk management of the models in light of the current threats. Security breaches or regulatory non-compliance events trigger external reporting to relevant supervisory authorities within statutory timelines.

The retrospective process produces corrective and preventive measures designed to prevent incident recurrence. Measures are integrated as applicable to the updates of Malt's AI system at stake, Malt's security, and business continuity policies.

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

Yes, both receiving and sharing

If yes, please indicate which stakeholders reporting mechanisms may be accessible to, recognising specifics will depend on the nature of the report and risk level. For each, indicate whether access is on a standing basis (Always), activated based on context (Ad hoc/as relevant), or not currently accessible.

	Always	Ad hoc / as relevant	Not currently accessible
Relevant employees (security, legal/compliance, etc.)	☐		
Public authorities (e.g., public reporting offices for security incidents or data breaches)		☐	
Supervisory authorities (e.g., data protection authorities)		☐	
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, etc.)			
Specific civil society organizations		☐	
Media / general public (people or communities indirectly impacted)			
Other			

Please describe how your organisation determines when and how reporting mechanisms are made accessible to different stakeholders (e.g. based on legal or regulatory requirements, internal policies, or risk level). Please also provide public URLs or details on how to submit reports. If none, write “Not applicable”

As of today, Malt's reports are available to stakeholders based on legal requirements.

Vulnerability reports can be submitted through Malt's bug bounty program, following this [link](#). Scientific reports, including on severe risks, can be found on Malt's dedicated [research page](#).

We further encourage users to file any report regarding AI safety and transparency either through the Malt interface or by sending an email to compliance@malt.com.

How does your organization share relevant information as appropriate?

- Reporting to regulatory authorities as relevant
- Publicly
- Information-sharing partnerships (with confidentiality agreements as appropriate)

Please describe what type of information is shared.

Technical information is shared with research partners for the purposes of performing research, and will be shared publicly through scientific papers (more information [here](#)). Statutory information is shared with regulatory authorities where required.

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

- Paid rewards (“bug bounty”)
- Research grants, contracts, and hiring red teamers
- Operational incentives for customers and partners
- Early access and collaboration incentives

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

Alongside non-incentivized initiatives like dedicated communication channels to encourage feedback reporting by employees when a new model is released, or a new product feature integrating AI; Malt operates an annual pen-test exercise, that includes red teaming, and a bug bounty program to identify and remediate potential security vulnerabilities. More information can be found [here](#).

Further, Malt has a dedicated in-house research team that collaborates with over four international university research centers on empirical studies and industry research to improve hiring in the labor market. Scientific papers published can be found [here](#).

In addition, Malt has stand alone incentivized field experiments by making users participate to analyze project briefs by freelancers alongside reviews of anonymized profiles by clients against compensation. Upon release of a new model, or a new product feature that involves AI, users are also provided with early access (no user-centric incentivized initiative ongoing at the moment).

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)
- Industry-led best practices or codes of conduct

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.

Malt contributes to the development of industry tools, evaluation frameworks, and technical best practices for AI risk assessment—particularly focusing on the safe use of AI in recruitment, hiring, and the freelance work domain.

We collaborate with the broader scientific and engineering community to build standardized tools for assessing AI risks and performance. We share our findings publicly through our [research page](#) and [engineering blog](#). For example, Malt researchers recently co-authored and contributed to establishing a standardized, community-driven open-source [benchmark](#) for AI in the work domain.

We further participate in the [SEISMEC](#) international consortium, where Malt’s pilot focuses on a core issue for digital labour platforms: matching needs to talent efficiently, while ensuring that decisions do not create systematic differences in visibility and access to opportunities between groups of freelancers.

The pilot develops and integrates evaluation tools that assess matching systems throughout their lifecycle.

Last, we participate in technical and regulatory working groups, and we contribute frequently to technical and regulatory panels.

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).

Malt distinguishes between recommendation systems, generative AI and agentic AI. Generative systems receive additional controls for prompt and output security, factuality, privacy, provider processing and human review. Agentic systems, where applicable, require additional safeguards such as least-privilege access, approved tool lists, action limits, human confirmation for consequential actions and an effective interruption or shutdown mechanism.

The lifecycle process is reviewed and updated under three circumstances: (1) when a new system version or model iteration is released, including consultation with the AI Deployment Committee for new algorithmic model baselines; (2) at least annually, including extensive independent penetration testing and a broader review of the effectiveness of governance, security, privacy, fairness and monitoring controls; (3) on an ad hoc basis, following an active model defect, vulnerability, operational exploit, material incident, significant performance or fairness degradation, major data or architecture change, new third-party dependency, regulatory development or material research finding.

For the plan and design stage, before developing or deploying an AI system, Malt defines in the design specifications its intended purpose, users, affected stakeholders, expected benefits, foreseeable misuse and potential risks. Systems are assessed in the risk mapping across privacy, security, fairness, transparency, accuracy, human rights, operational resilience and environmental considerations by the AI deployment committee. The review determines whether the system can proceed, requires additional safeguards or should not be deployed.

In the collect and process data stage, Malt uses professional and marketplace information relevant to the service, such as skills and experience for matching and search. Malt evaluates data quality, completeness, and representativeness. Data is limited, cleaned, and tested, including for potential bias, before use. Users can correct, update, or request deletion of their information. Third-party models and AI services are subject to AI, privacy, security and contractual review. Data shared with external providers is governed by controls such as confidentiality, zero-retention policy, limited access.

In the build, integrate, and interpret model stage, Malt's matching systems use layered architectures combining lexical and semantic retrieval, ranking and business rules. Models, prompts, datasets, configurations and third-party dependencies are documented and versioned where relevant.

Development and testing are separated from production environments, and access is controlled according to role and need. User-facing explanations focus on the purpose and limitations of AI-assisted search and recommendations rather than presenting the system as an autonomous decision-maker. Proposed candidates, and AI-generated briefs remain subject to human validation. Users can edit, reject or replace AI-generated recommendations and can use keyword search in the directory and type in manually their brief.

Through the verify and validate stage, Malt verifies AI systems before and after material changes through a combination of automated testing, offline/online evaluation, human review, comparative experiments, security testing and fairness analysis. Malt also evaluates user group performance and potential disparities by including group analysis and counterfactual testing in which relevant characteristics are changed while professional skills and experience are held constant. These evaluations help identify differences in ranking position, visibility and exposure.

Before release, at the deploy stage, Malt considers the evaluation results. Where appropriate, changes are introduced progressively and monitored against an approved baseline. Production systems maintain logging and traceability sufficient to investigate outputs, model versions, relevant inputs and human interventions, while respecting data-minimization requirements. Malt maintains human oversight for AI-assisted matching and recommendations. Fallback and rollback procedures are designed to allow Malt to isolate a faulty component, revert to a previously validated version or use a deterministic alternative when an AI component or external service is unavailable or behaving unexpectedly.

While in operation, Malt monitors AI systems across data quality, performance, fairness, security, reliability, user feedback and operational impact. Monitoring may include error rates, latency, anomalous behavior, user complaints, and external-provider failures. AI-related incidents may result in additional testing, changes to data or prompts, enhanced human review, rollback, temporary suspension, retraining or retirement of the affected system. Feedback from clients, freelancers, employees, researchers and other stakeholders contributes to continuous improvement.

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).

Malt's proprietary matching models are developed using first-party data generated on the Malt platform and focusing on professional information relevant to the intended purpose, such as skill and expertise. The data is collected in the context of providing and improving Malt's services. Malt does not rely on public-web scraping for the training of its proprietary matching models.

Malt applies data-quality and data-governance measures throughout collection, transformation and curation. These include documented data source, controlled access, schema and consistency checks, moderation of inappropriate content, validation of professional attributes, monitoring of data freshness. Data pipelines and transformations are subject to quality checks and lineage controls so that material changes can be identified and investigated. In addition, non-AI business rules govern specific operational parameters, including candidate location, availability, and language skills. Where lawful and necessary, protected and aggregated sensitive personal data may be used separately for fairness testing and impact assessment. Data minimization, retention controls and access restrictions are applied to derived datasets, evaluation data and other model-related artefacts.

To reduce the risk that historical booking patterns reproduce existing biases, Malt evaluates the quality and fairness of training and evaluation data through group analysis, counterfactual testing and audits. This includes testing whether profiles with comparable skills and experience receive different ranking positions when protected characteristics are varied. Data-quality, fairness and representativeness findings are incorporated into model development, deployment reviews and post-deployment monitoring. Material deterioration, evidence of disparate impact, data-quality failures, harmful outputs or changes in the underlying data can trigger additional investigation, corrective action, retraining, rollback or a renewed review by the relevant governance and deployment functions.

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

- Isolated and secure test environments
- Isolation of staging environments (from production / operational environments)
- Controlled environments for red-teaming and adversarial testing
- Controlled security vulnerability testing
- Rate limits and resource caps
- 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.

The information security team is in charge of controlled environments for red-teaming and adversarial testing, controlled security vulnerability testing, and access controls, including role-based permissions.

The machine learning team is in charge of isolated and secure test environments, isolation of staging environments, use of “AI-safe” test-data, as appropriate, the use of synthetic, anonymized, or fake data, as appropriate, and human-in-the-loop controls for higher-risk test cases.

The data platform team is in charge of logging and monitoring.

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:

- Not applicable

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

- Not applicable

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.

Malt applies the fundamental data-protection principles under the GDPR throughout the collection, processing, storage and deletion of personal data.

Personal data used in Malt’s matching and search services is processed for specific purposes, including providing freelancers with relevant project opportunities and helping clients identify freelancers whose professional skills and availability correspond to their project requirements. Malt limits processing to data relevant to these purposes and applies additional safeguards to data used for model development, evaluation and monitoring.

Malt’s privacy governance in the context of the AI system is organized around two main areas. First, privacy by design is integrated into product and AI development, the DPO is consulted before a new model version or material model change is initiated so that the relevant data protection impact assessment can be performed. Second, the upholding of individual's rights throughout the lifecycle stages of the AI system. Freelancers and clients can access, update and correct relevant professional information through their accounts or through dedicated privacy channels. More information is available in Malt’s [privacy policy](#).

Data protection standards align with ISO27001:2022 requirements (see below).

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)
- Implement incident response, disaster recovery, or business continuity plans, including escalation procedures for high-severity incidents
- Train staff on cybersecurity best practices and AI-specific threats, including how to recognize and respond to misuse or unintended outcomes of advanced AI systems
- Regularly review security measures and update them in response to emerging risks, incidents, or major releases

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
- Restrict physical access to servers, GPUs, or devices hosting AI models, systems, and/or applications
- Monitor system activity and log access to detect anomalies or misuse
- Conduct secure AI development and/or deployment practices
- Conduct security testing focused on AI-specific risks (e.g., red-teaming, adversarial testing, stress tests)

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
- Track and log access to sensitive AI resources
- Require NDAs or confidentiality agreements for staff and contractors with access
- These responsibilities are delegated contractually to model or infrastructure providers

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

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

Malt integrates AI cybersecurity into its wider information-security framework rather than treating it as a separate control environment. Malt applies a security framework aligned with ISO/IEC 27001 standards. Malt ensures formalized risk management, continuous security objectives tracking, and regular internal compliance reviews. System access is governed by role-based access control, least privilege principles, and mandatory multi-factor authentication across all corporate systems, developer repositories, and production environments. All platform data is encrypted in transit using strong cryptographic protocols

(TLS 1.3) and at rest using industry-standard encryption (AES-256) across databases and file stores. Malt undergoes extensive, independent third-party penetration testing on an annual cycle to proactively identify and remediate code vulnerabilities and configuration flaws. And Malt continuously monitors its information system, including through a bug-bounty program.

Malt classifies models, databases, prompt templates, and inputs as high-sensitivity assets requiring specialized technical controls such as filters to remove unneeded personal identifiers and confidential information; further, the information system runs on dedicated, private cloud server capacity with restricted network access controls; client project briefs, and user prompts processed via third-party generative LLMs (e.g., Gemini family) are subject to strict zero data retention policies.

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
- Bug bounty or incentive programs for external reporters
- Investigation of reported or identified risks and vulnerabilities
- Development of patches or other mitigations for confirmed risks or vulnerabilities
- Testing of patches or other mitigations
- Scanning for and mitigation of similar risks or vulnerabilities
- Defined escalation criteria and follow-on actions for identified vulnerabilities (e.g., for external reporting or information sharing)
- Reporting to regulators / authorities when relevant

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

- 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)
- Perform root-cause analysis and apply lessons learned to improve future development and/or deployment practices
- Conduct post-deployment monitoring

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)

- Monitoring (anomaly detection, multi-agent interaction monitoring)
- Logging of agent actions

Please provide any further comments and implementation documentation.

Malt's agentic system requires additional safeguards such as least-privilege access, approved tool lists, action limits, detailed trace logging, human confirmation for consequential actions and an effective interruption or shutdown mechanism.

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?

- 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).
- Technical documentation and instructions for use if relevant.
- Plain-language or non-technical summaries

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

Malt publishes public-facing information about its AI-assisted search through product explanation available on Malt's dedicated [AI page](#). Further information can be found in Malt's [privacy](#) and [transparency](#) policies. On an ad hoc basis, Malt will do webinars, or articles that share further information (see examples [here](#)). These materials describe, in accessible language: the purpose and main capabilities of the AI systems; how AI assists search, matching, recommendations and, where applicable, the preparation of project briefs; the types of information used, at an appropriate level of detail; the appropriate use of the systems as assistance tools; inappropriate uses, including treating AI outputs as final hiring or contracting decisions, or relying on outputs without human review; the role of human oversight, user control, and available correction or complaint channels; and the way Malt evaluates, monitors and updates its AI systems.

The public information is written for non-technical audiences and is complemented, where appropriate, by more detailed technical and compliance documentation for customers, auditors, regulators and other qualified stakeholders in compliance with the regulation.

Publicly available technical information is available on our [research page](#) and [engineering blog](#).

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
- Protected information sharing
- Regulatory reporting as relevant
- Direct engagement with AI safety institutes, civil society or academic partners

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.

Malt has a dedicated [AI page](#) whereby Malt shares in non-technical, plain-language summaries of the description of Malt's AI system and how it works, as well as our principles guiding AI developments, and security, privacy, and AI safeguards.

Detailed information, including evaluation results may be shared confidentially with customers, enterprise partners, auditors, regulators, research collaborators or other qualified stakeholders where this is necessary to demonstrate responsible use or support risk mitigation. Depending on the recipient and purpose, this may include technical documentation, evaluation reports, fairness analyses, security-assessment findings, model limitations, monitoring results and remediation plans.

Malt provides information to competent authorities where required by law or where regulatory engagement is appropriate. Regulatory information may include system descriptions, data-flow information, risk and impact assessments, testing results, incident information, corrective actions and relevant governance records.

Where appropriate, Malt also engages with public institutions, civil-society organizations, industry groups and AI-safety or responsible-AI initiatives. These collaborations support the assessment of matching quality, fairness, bias, explainability and the broader socio-economic effects of AI-assisted access to freelance opportunities. An example can be found [here](#).

Malt shares research findings through peer-reviewed publications, conference presentations, technical or engineering communications, product and policy documentation. These materials may describe evaluation methods, performance results, fairness and bias findings, known limitations, security and privacy safeguards, human-oversight arrangements and mitigation measures. See our research page [here](#).

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

Yes

If yes, provide links to privacy policies or notices covering personal data, user inputs, and AI-generated outputs.

Malt maintains comprehensive, publicly accessible [terms and conditions](#) of the platform, and [privacy policy](#) that explicitly governs the collection, processing, and protection of personal data, user inputs, and system-generated outputs across the platform. Additional information can be found in the dedicated [AI page](#).

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?

- Protected information sharing
- Regulatory reporting as relevant
- Direct engagement with civil society or academic partners
- Public documentation

Please provide links to any publicly available information.

Malt does not train its AI system on external datasets - the source of the data used for the training of the AI system is Malt's platform. The source of data used for the training is detailed in Malt's publicly accessible [privacy policy](#) that explicitly governs the collection, processing, and protection of personal data. The [terms and conditions](#) provide for the use of the user inputs, and system-generated outputs across its platform. Additional information is available on Malt's dedicated [AI page](#). Further information about specific training datasets can be shared confidentially through protected information sharing, regulatory reporting, and direct engagement with civil society or academic partners.

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).

Malt analyzes and publishes anonymized, aggregated market data to inform broader economic understanding regarding freelance labor dynamics, emerging skill demand, and digital workforce transformation across sectors and European markets. See the latest Malt tech trends report [here](#).

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
- Senior management oversight
- Governance council or committee
- Board of Directors 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).

The governance structure distinguishes between advisory, operational and final decision-making responsibilities. System documentation and assessments are reviewed when a new model or system version is released; and ad hoc reviews are triggered by a major incident, a material change to the model, prompt, data, architecture, provider or intended use, a significant deterioration in accuracy, fairness, security or user outcomes.

Malt's Board has an advisory role. It assesses and advises on issues such as algorithmic fairness, and may recommend additional safeguards. Malt also involves a Freelancer Advisory Board of seventeen freelancers selected to reflect diversity in expertise, geography and socio-economic background. The board is consulted annually on progress, ethical considerations and emerging trade-offs between business objectives and fairness, ensuring that those affected by platform decisions have a formal channel into governance.

The AI Deployment Committee has operational decision-making authority for AI deployments and material changes. It reviews the relevant risk classification, impact assessments, data and model documentation, evaluation results, security and privacy controls, human-oversight arrangements, monitoring plans and fallback or rollback procedures. It may approve a system or model version, approve it subject to conditions and remediation, require additional testing, reject deployment or suspend an existing system.

Product, data and machine-learning engineers (owners) are responsible for implementation. They operate the system, maintain the technical documentation, implement approved controls, monitor performance and fairness, investigate anomalies and ensure that remediation actions are completed.

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).

New employees complete AI, data protection, and security trainings during onboarding. This includes an introduction to responsible AI, data-protection principles, information security, acceptable use of AI tools, human oversight and the organisation's corporate AI policy. Employees are required to review and formally acknowledge the relevant policies.

Malt also provides continuous and role-based targeted workshops, practical guidance, policy updates and refresher sessions. The content is adapted for technical, product, data, security, legal, privacy and operational teams. Depending on the role, training may cover: AI governance and risk classification; privacy by design and data minimization; fairness, bias and responsible use of professional data; secure development and protection of AI assets; model, prompt and data lifecycle management; incident response, vulnerability reporting and escalation; appropriate and inappropriate uses of generative and agentic AI.

Employees can raise concerns through the "Speak Up" reporting channel. This channel can be used to report suspected misuse, discriminatory outcomes, privacy concerns, security vulnerabilities, unsafe model behavior or breaches of Malt's AI policy. The reporting channel may provide for anonymity [here](#). Reports are handled according to the applicable escalation and incident-management procedures, with access limited to the compliance and security teams.

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

Yes

If yes, describe how incident records are kept and used for follow-up (e.g., incident logs, root-cause analysis reports, internal tracking systems). You may indicate whether and how you define escalation thresholds for when incidents must be surfaced to senior leadership or external parties.

Malt manages AI-related incidents through its general incident management policy, which is integrated with Malt's information security management system (ie. tracking system), the information security policy and the data breach policy. Every incident is recorded in a live incident document which includes a root cause analysis report. Once the incident is resolved, a post-incident analysis is conducted to further understand root causes, control gaps and measures required to prevent recurrence.

Escalation thresholds are based on severity and impact. Escalation may involve the system owner, ML or engineering teams, information security team, the DPO, legal, senior management. A critical impact on users, legal obligations, finances, productivity or reputation may trigger escalation to the executive committee. Incidents are shared with external parties—such as regulatory authorities and impacted users—when legally mandated or to prevent unintended outcomes.

Incident records and the personal-data breach register are used for audits, accountability, trend analysis, risk-register updates, control improvements, training and future deployment decisions. Major incidents, material system changes, new vulnerabilities, significant evaluation findings and regulatory

developments trigger an ad hoc review of the relevant AI risk assessment and controls by the AI deployment committee.

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).

Malt participates in research collaborations with academic and external research [communities](#); and Malt engages in public body consultation calls such as the consultation on the [draft guidelines](#) for the classification of high-risk artificial intelligence systems, technical or regulatory panels, regulatory working groups. These collaborations support independent evaluation, the development of common benchmarks and methods, and the assessment of broader effects of AI on work, matching and access to professional opportunities.

Further, Malt research scientists share findings through conferences, workshops, lectures and other scientific or technical forums (example can be found [here](#) and [here](#)). Malt shares its research and good practices through several channels; it publishes papers at world renown conferences (RecSys) and on its [research website](#), including work on AI evaluation in the work domain, and fairness across groups. In addition, Malt's [engineering blog](#) provides practical explanations of how its technology, data and product teams design, implement, evaluate and operate AI-enabled systems. These publications share technical approaches, architecture decisions, performance results, implementation lessons and operational considerations with the wider engineering and research community.

Malt shares sufficient methodological and contextual information to support meaningful review while protecting personal data, confidential information, security-sensitive details and proprietary assets. This enables external review, discussion and comparison with approaches developed by other organizations.

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:

- ISO/IEC standards (e.g., ISO/IEC 27001, ISO/IEC 42001)
- NIST AI Risk Management Framework (AI RMF)
- Industry-led best practices or codes of conduct
- Other international or regional standards (please specify): EU AI Act/GDPR

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

Malt uses a combination of technical standards, regulatory frameworks and industry best practices to structure AI governance framework.

The AI Act and GDPR provide AI-system classification, risk and impact assessments, data governance, privacy and fundamental rights by design, human oversight, transparency, technical documentation, monitoring, and individual rights.

ISO 27001 provides the foundation for Malt's information-security management system which covers for example access control, encryption, vulnerability management, secure development, supplier due diligence, incident response, business continuity and continuous improvement.

The risk management framework is a combination of traditional risk-mapping that integrates the NIST AI Risk Management Framework, particularly the AI lifecycle stages and trustworthy AI characteristics. The technical risks are addressed through industry-led, and technical research works on model and data evaluation, fairness testing.

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)
- Documentation or FAQs explaining when and how AI is used in the system
- Disclosure to end-users when an AI system is acting autonomously or executing actions on their behalf (e.g., making purchases, sending communications, modifying files)

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).

Malt informs users when they are interacting with AI through product labelling (see for example the "[AI Search](#)" product), user-interface information through the [AI page](#), and dedicated disclaimers when they are using the conversational AI where applicable. This informs users that AI is being used to assist the interpretation of project requirements, the retrieval and recommendation of freelancer profiles and, where applicable, the preparation of project information.

Malt's AI documentation explains the purpose of the AI features, the types of assistance they provide, their main limitations and the continuing role of the user. Users are informed that AI outputs are recommendations or drafts and should be reviewed. They can modify search criteria, edit or reject generated content, use manual search and challenge or report an inappropriate result.

The disclosure mechanisms are reviewed when new AI features or model versions are introduced, when the system's degree of autonomy changes, or when user feedback, incidents or regulatory developments indicate that additional transparency is required.

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

Not applicable

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
- System performance and outcomes (e.g. accuracy, consistency)
- Explainability or interpretability
- Transparency
- Robustness or reliability

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.

Malt invests in research and technical development to improve the safety, security, fairness, transparency, explainability, reliability of AI systems used in the work and freelance domain. Malt's research team includes in-house researchers with academic expertise (see more information [here](#)). Malt also collaborates with external academic and research communities. This includes participation in the EU-funded [SEISMEC consortium](#), where the pilot develops and integrates evaluation tools that assess matching systems throughout their lifecycle. This makes risks visible early and supports evidence-based go or no-go decisions before changes are deployed at scale.

Key areas of research and investment include: (i) data quality and representativeness by studying the quality, coverage and limitations of data used for professional matching and ranking. This includes work on multilingual data, profile and project representation, historical selection effects and the risks of using marketplace outcomes as training or evaluation signals (eg. [An Efficient Long-Context Ranking Architecture With Calibrated LLM Distillation: Application to Person-Job Fit](#)); (ii) fairness, safety and responsible use by investigating algorithmic bias and the potential effects of AI-assisted matching on access to professional opportunities. Research includes the evaluation of implicit preferences, group disparities, counterfactual profiles and the possible reproduction of historical or user-driven biases (eg. [Evaluating LLM Behavior in Hiring: Implicit Weights, Fairness Across Groups, and Alignment with Human Preferences](#)); (iii) explainability and interpretability by reviewing how matching and ranking systems can be understood and evaluated, including the factors influencing relevance scores, model

behavior and ranking outcomes. This supports technical diagnostics, user explanations and the design of appropriate human oversight (eg. [WorkRB: A Community-Driven Evaluation Framework for AI in the Work Domain.](#)).

Research outputs are shared publicly through papers, technical articles and conferences. More detailed evaluation results, datasets or security information may be shared confidentially with research partners, customers, auditors or regulators where appropriate. Some information remains internal where disclosure could create privacy, security, intellectual-property or competitive risks.

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

- Socio-economic considerations (e.g., AI adoption and productivity, workforce transitions, skills development, access to AI tools)
- 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.

Malt has a strong corporate social responsibility culture since day one (more information can be found [here](#)). Malt researches and invests in methods intended to reduce the socio-economic and environmental risks associated with AI-assisted matching.

With regards to socio-economic risks, Malt's research examines how algorithmic matching can affect access to professional opportunities, and the distribution of work. This includes research on user preferences, group disparities, and the potential for recommendation systems to reproduce existing inequalities (see above).

With regards to environmental considerations, Malt follows a task-appropriate model-selection approach. Large generative models are not used where a smaller, specialized model can perform the task adequately (see for example this [tech conference talk](#)). Malt continuously strengthens its work by measuring energy consumption and carbon impact across material AI workloads.

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.

Malt promotes AI literacy through a combination of internal training, user education and public-facing information.

For employees, the Malt AI training path provides onboarding-level training on AI governance, responsible use of AI, privacy, information security and the organization's AI policy. The AI training path also provides for role-based learning for technical, product, data, security, legal and operational teams. For freelancers and clients, Malt provides in partnership with continuous learning institutions specialized trainings "AI for all" where users are invited to automatize, analyze and innovate through AI (see [this series of webinars](#) for example). Furthermore, Malt hosts dedicated resources such as "[AI for freelancers](#)" to provide guidance/tips on AI literacy and "tips", as well as "[AI projects](#)" showcasing success stories.

Malt also shares public research and technical explanations through its [research website](#) and [engineering blog](#). These materials provide broader information on AI evaluation, fairness, matching, model behavior and the potential impact of AI on work and professional opportunities.

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.

Malt pursues AI research and product initiatives that aim to generate broader societal benefits, particularly by supporting fairer access to professional opportunities, reducing bias in digital matching and improving transparency in AI-assisted work platforms. Examples include the [Malt Trends Seniors](#) report. These initiatives are intended to support more inclusive, transparent and accountable digital labor markets.

Any further comments and implementation documentation.

No answer provided