

OECD AI Transparency Report

Organization: ARTEFACT (FRA) 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)
- Deployer

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

- Maintain an internal taxonomy, classification system, and/or set of risk area definitions
- 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): NIST AI RMF, ISO/IEC 42001, OECD AI Principles

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.

Artefact defines thresholds for unreasonable risk primarily through our Internal AI Usage Guidelines, which set out a list of strictly prohibited AI uses on legal and ethical grounds, aligned with the EU AI Act. These include: social scoring, mass surveillance, workplace emotion recognition, manipulative AI, and predictive policing. Any use case falling into these categories is considered to present an unreasonable level of risk and is not permitted, regardless of business justification.

Also, the data consulting security template has a dedicated "AI Management" section (AI-specific risk assessment, AI Act compliance, human oversight, transparency) → AI risk assessment is built into the standard project process.

In addition, Artefact applies a company-wide severity-scoring methodology, developed as part of our CSRD double materiality analysis, which rates risks according to magnitude, scope, irreversibility, and likelihood on a 1–5 scale. While this methodology was originally designed for financial and impact materiality assessment rather than AI-project risk specifically, it reflects our broader organizational practice of applying structured, quantified risk assessment rather than relying solely on qualitative judgment.

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 benchmarks, standard metrics (e.g., frequency, severity scores)
- Qualitative assessments (e.g., expert judgment, narrative scoring)
- Impact assessments such as personal data protection assessment and fundamental rights impact assessments
- Specific risk identification practices for agentic AI capabilities, e.g., autonomous action-taking, tool use, multi-step planning (please specify): For example Agentic supervision study

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.

Artefact applies a Data Science Project Checklist across its AI projects, which includes a dedicated pre-deployment review stage. Before any model or system moves into production, this checklist requires: testing in a production-like environment, external review by domain experts, a fairness review, an assessment of potential misuse or harmful side-effects, a compliance check against applicable regulations (e.g., GDPR), and the definition of a fallback strategy in case of failure or unexpected behavior.

In addition, each project is governed by a dedicated steering committee and requires formal validation from both the CISO (RSSI) and the Data Protection Officer (DPO) before deployment can proceed.

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.

Incidents reports are used following the guidelines listed in our incident management policy.

Artefact relies on formal email communication channels and dedicated points of contact to receive and share reports regarding risks, vulnerabilities, and incidents.

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”

This is determined according to internal policies.

How does your organization share relevant information as appropriate?

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

Please describe what type of information is shared.

This is determined according to our security policy.

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

- Internal incentives and culture

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

Details are not publicly accessible.

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?

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			☐

	Always-on	Ad hoc / as relevant	No current collaboration
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.

We collaborated for example through the Positive AI initiative.

Artefact participates in Positive AI, a multi-sectoral initiative focused on ethical and responsible AI, which we joined in July with the explicit goal of collaborating with other organizations across sectors. Given the recent nature of this membership, our participation is currently ad hoc rather than standing, as the initiative has not yet been fully rolled out internally.

On the academic side, Artefact maintains active research partnerships, notably with University Toulouse 3 and INRIA, focused on statistical models for AI risk assessment and mitigation. These collaborations produce concrete outputs, including published papers and open-source Python packages implementing these methods.

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).

Artefact manages AI risk across the full project lifecycle through its Data Science Project Checklist, which structures every AI project around six stages: problem framing and business justification (Plan and design), data access and quality assessment (Collect and process data), baseline definition and model development (Build, integrate and interpret model), performance, bias and reproducibility checks (Verify and validate), production-readiness review including external expert review and regulatory compliance (Deploy), and ongoing metric tracking, drift detection and rollback capability (Operate and monitor).

For generative AI use cases specifically, this is complemented by our GenAI Risk Strategy, which maps concrete risks and remediation plans, both provider-side and internal, for each client use case (e.g., data exposure, IP, hallucination, GDPR compliance).

For each agentic AI system deployment, we apply a dedicated governance framework covering the full agent lifecycle from ideation to build and run, with defined roles and responsibilities across business, technical and control functions (security, data protection, legal), and specific oversight of agent autonomy, monitoring and emergency intervention.

Reviews are triggered by project milestones (go-live decisions), by incidents, and by material changes to a deployed system; cross-cutting escalation is coordinated through our "AI Security Office". We have an AI Security Office and we have an ISP Policy (Integration of security into projects) and procedure.

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).

Data quality and fairness are addressed at multiple stages of our project lifecycle. During data collection and preparation, our Data Science Project Checklist requires an assessment of whether the available data is representative of the population of interest, and whether minority groups are sufficiently represented to avoid biased outcomes. Before deployment, models undergo a dedicated fairness review, including manual inspection of misclassified examples and analysis across sensitive attributes such as race and gender.

This is reinforced by dedicated open-source research packages developed and published by Artefact: OT-FairBoost, which improves the trade-off between model accuracy and reduced discrimination (developed in partnership with L'Oréal, University Toulouse 3, CNRS and INRIA); FairGroups, a discrimination-detection tool presented at ECAI; and Woodtapper, which enables auditing of tree-based models through simple, human-interpretable rule sets (developed in partnership with Société Générale).

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

- Isolated and secure test environments
- Network isolation from external access
- Isolation of staging environments (from production / operational environments)
- Controlled environments for red-teaming and adversarial testing
- Containment measures for prompt-injection and jailbreak attempts
- Controlled security vulnerability testing
- 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:

- Publishing “model cards” or contextual information on training data and model limitations
- 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
- Conduct data cleaning, qualification, and/or tokenization activities as part of a transformative data preparation process
- Consult and/or retain an intellectual property attorney to protect, enforce, and comply with, intellectual property rights.

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
- Other(s) (specify): mandatory use of specific AI tools (internal, secure) for client/sensitive data; secure API keys (Azure OpenAI) instead of public tools; prohibition on uploading client documents to public AI

Describe key privacy measures and data-protection standards applied.

Artefact protects privacy and prevents disclosure of confidential or sensitive data through internal governance embedded in our Internal AI Usage Guidelines. Employees are required to use Gemini, our internal and secured GPT environment, whenever client data, sensitive information, or contractual commitments are involved, rather than public AI tools. Where external AI tools are used, employees must rely on secure API keys (e.g., Azure OpenAI) rather than free public versions.

Uploading client documents, deliverable templates, or sensitive files to public or unsecured AI platforms is explicitly and strictly prohibited. Employees are also required to review the terms of use of any AI tool before adoption, with particular attention to clauses on data ownership, content reuse for model training, and confidentiality, supported by internal guidance identifying red-flag contract language (e.g., broad content licenses, lack of confidentiality guarantees).

Any doubts regarding a tool or practice are escalated to our dedicated AI Security Office. In addition, each project maintains a personal data processing log covering purposes, legal basis, and applied security measures, which requires formal validation from both the CISO (RSSI) and the Data Protection Officer (DPO).

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
- Integrate cybersecurity AI risks into broader cybersecurity risk management activities

- Regularly review security measures and update them in response to emerging risks, incidents, or major releases
- These responsibilities are delegated contractually to model or infrastructure providers

11.B. Technical and operational controls

- Apply multi-factor authentication (MFA) and role-based access controls for advanced AI systems
- Conduct secure AI development and/or deployment practices
- Conduct security testing focused on AI-specific risks (e.g., red-teaming, adversarial testing, stress tests)
- These responsibilities are delegated contractually to model or infrastructure providers

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

- Limit access to proprietary AI assets based on role or need-to-know
- Track and log access to sensitive AI resources

11.D. Insider risk protection detections

- 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.

Artefact's approach to AI security is anchored in a dedicated internal function, the AI Security Office, which serves as the central point of contact for any question or concern regarding AI tool usage or security practices across the organization. This function maintains and regularly reviews a governance mechanism for AI tools, including an approved-tools framework: a blacklist of prohibited AI tools and a whitelist of approved transcription tools, both subject to periodic review, with employees responsible for checking approval status before using any tool not already listed.

In addition, every client project is governed by a dedicated security template, validated by the CISO (RSSI) and, where personal data is involved, the Data Protection Officer (DPO), under the oversight of a project steering committee. For technical/data consulting engagements, this includes scheduled penetration testing, security audits, attack simulations, and coverage of the OWASP Top Ten vulnerabilities. At contract close, a formal process governs asset restitution and access revocation. Non-compliance with these security rules can result in disciplinary sanctions under Artefact's AI Usage Charter, which carries the same legal standing as internal company rules.

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
- Testing of patches or other mitigations
- Defined escalation criteria and follow-on actions for identified vulnerabilities (e.g., for external reporting or information sharing)

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

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?

Not applicable

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?

- Protected information sharing

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.

Artefact shares evaluation, risk, and impact assessment results across three primary stakeholder groups :

- **Enterprise Clients** : For client deployments, technical evaluation results, red-teaming outputs, and continuous performance metrics (such as RAGAS scores, hallucination rates, and security vulnerability scans) are shared confidentially. This is conducted through dedicated project steering committees, formal audit reports, and customized observability dashboards (e.g., using tools like Langfuse)
- **Research Community** : We collaborate directly with academic institutions and research bodies (including Sorbonne University, Toulouse 3, CNRS, and INRIA) to share methodologies and evaluation tools. Findings on model interpretability, bias mitigation, and safety are published in peer-reviewed scientific conferences (e.g., AISTATS, ECML-PKDD, ECAI) and released as open-source auditing packages (e.g., Woodtapper for model auditing, OT-FairBoost for fairness, and MGS-GRF).
- **Industries** : We engage with the broader ecosystem through active participation in multi-stakeholder initiatives such as Positive AI, the Coalition for Sustainable AI, and the UN Global Compact. Public insights, white papers (e.g., The Future of Agentic Supervision), research newsletters, and thought leadership articles are regularly published to advance collective industry standards around responsible AI adoption.

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

Not applicable

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?

- Not applicable

Please provide links to any publicly available information.

Not Applicable

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

Yes

If yes, describe additional method(s), including where relevant transparency measures related to AI agents

No answer provided

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?

Not applicable

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?

- Governance council or committee

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 at Artefact is embedded through a two-layer governance structure. Day-to-day AI tool usage and security questions are centralized through a dedicated internal function, the AI Security Office. At the project level, every engagement is governed by a dedicated steering committee, with formal validation required from the CISO (RSSI) and, where personal data is involved, the Data Protection Officer (DPO), before deployment can proceed. Non-compliance with these governance requirements can result in disciplinary sanctions under Artefact's AI Usage Charter, which carries the same legal standing as internal company rules.

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).

Specific trainings are deployed.

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.

Incidents are tracked internally.

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).

Research Center publications, Agentic Supervision study, blog articles

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:

- NIST AI Risk Management Framework (AI RMF)

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

ISO/IEC 27001

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?

- 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).

Artefact's approach to client project delivery includes the principle that end-users should be informed when interacting with an AI system, and specifically when an AI agent is acting autonomously on their behalf. Implementation of this principle is determined on a project-by-project basis in agreement with each client, rather than through a single company-wide standardized mechanism.

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?

- System performance and outcomes (e.g. accuracy, consistency)
- Explainability or interpretability
- Transparency
- Other (please specify): Fairness & bias

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.

We have conducted a research project in partnership with Sorbonne University during 24 months to improve accuracy of AI models for scoring and classification, with practical target of imbalance data. We released an open source package, MGS-GRF, and published two papers, in AISTATS and ECML-PKDD, top conferences in machine learning, showcasing how our methods reach unprecedented results on many data-sets and problem of scoring and classification.

We have conducted several research projects to better interpret AI models. We developed a foundation model for retail, that is interpretable and reaches state-of-the-art accuracy. We also developed an open source package, Woodtapper, who has a module that provides an explanation of predictions by the example.

We have developed and released an open source package, Woodtapper, that enables to audit any tree-based model, the state of the art of machine learning for tabular data. It was in partnership with Sorbonne University and Societe General, a major French bank, who needs to audit their sensitive use

cases. The package enables to explain any model as a simple list of rules, easily interpretable by a human.

We have conducted several research projects to measure and improve the fairness of machine learning models. Our package, OT-FairBoost, reaches unprecedented trade-off between accuracy of model and reduced discrimination. We applied it to many scenarios, regression, classification, mitigation of disparate impact or equality of odds. It was conducted in partnership with l'Oréal, University Toulouse 3, and professors affiliated to CNRS and INRIA.

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)

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

We conducted for example research on AI and the Future of Work.

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.

Artefact supports AI literacy through a range of education and awareness initiatives designed to help clients, users and organisations better understand the capabilities, limitations and responsible use of advanced AI systems. These initiatives include AI training programmes, executive workshops, awareness sessions, conferences, publications and knowledge-sharing events such as Adopt AI. Through these activities, Artefact promotes informed AI adoption by encouraging a clear understanding of both the opportunities and the limitations of AI technologies, while fostering responsible and

trustworthy use.

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.

Artefact actively participates in multi-stakeholder initiatives that seek to advance societal and public benefits through the responsible development and deployment of AI. Through initiatives such as Positive AI and the Coalition for Sustainable AI, Artefact collaborates with industry peers, researchers and other stakeholders to exchange best practices, advance collective thinking on AI ethics, transparency and environmental sustainability, and contribute to the development and public dissemination of shared guidance and resources. These collaborative efforts aim to strengthen the broader AI ecosystem and support the development of trustworthy, responsible and sustainable AI for the benefit of society.

Any further comments and implementation documentation.

Artefact believes that advancing human and global interests requires collaboration across the AI ecosystem. Beyond implementing responsible AI practices internally, we actively contribute to multi-stakeholder initiatives that promote trustworthy, ethical and sustainable AI through knowledge sharing, collective action and the development of practical guidance. As a signatory of the UN Global Compact and an active participant in initiatives such as Positive AI and the Coalition for Sustainable AI, Artefact seeks to align its AI practices with broader societal and sustainability objectives. We remain committed to continuously strengthening our governance, sharing best practices and supporting the responsible development and deployment of AI for the benefit of society.