Scaling of End-To-End Governance Risk Assessments for AI Systems

Daniel Weimer
ceel.ai, Munich, Germany

Andreas Gensch
ceel.ai, Munich, Germany

Kilian Koller
ceel.ai, Munich, Germany

Abstract

Artificial Intelligence (AI) systems are embedded in a multifaceted environment characterized by intricate technical, legal, and organizational frameworks. To attain a comprehensive understanding of all AI-related risks, it is essential to evaluate both model-specific risks and those associated with the organizational and governance setups. We categorize these as "bottom-up risks" and "top-down risks," respectively. In this paper, we focus on the expansion and enhancement of a testing and auditing technology stack to identify and manage governance-related risks ("top-down"). These risks emerge from various dimensions, including internal development and decision-making processes, leadership structures, security setups, documentation practices, and more. For auditing governance related risk, we implement a traditional risk management framework and map it to the specifics of AI systems. Our end-to-end (from identification to monitoring) risk management kernel follows these implementation steps:

- Identify
- Collect
- Assess
- Comply
- Monitor

We demonstrate that scaling of such a risk auditing tool requires fundamental aspects. Those aspects include for instance a role-based approach, covering different roles in the development of complex AI systems. Ensuring compliance and secure record-keeping through audit-proof capabilities is also paramount. This ensures that the auditing technology can withstand scrutiny and maintain the integrity of records over time. Another critical aspect is the integrability of the auditing tool within existing risk management and governance infrastructures. This integration is essential to reduce the barriers for companies to comply with current regulatory requirements, such as the EU AI Act [3], and established standards like ISO 42001:2023. Ultimately, we demonstrate that this approach provides a robust technology stack for ensuring that AI systems are developed, utilized and supervised in a manner that is both compliant with regulatory standards and aligned with best practices in risk management and governance.

2012 ACM Subject Classification Computer systems organization

Keywords and phrases AI Governance, Risk Management, AI Assessment

Digital Object Identifier 10.4230/OASIcs.SAIA.2024.4

Category Practitioner Track

1 Motivation

The rapid adoption of AI systems across various industries has introduced significant challenges in managing and governing the associated risks. These risks, including algorithmic bias, data privacy concerns, and security vulnerabilities, demand comprehensive risk management frameworks that ensure safety, fairness, and regulatory compliance. However, the increasing

© Daniel Weimer, Andreas Gensch, and Kilian Koller;
licensed under Creative Commons License CC-BY 4.0

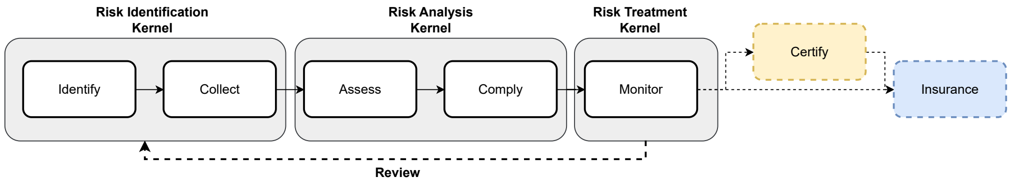

Figure 1 End-to-end workflow for AI risk management. The circular workflow is based on the general standard for risk management, described in [2].

complexity and scale of AI systems make manual risk assessments insufficient to effectively address these issues, necessitating automated solutions for timely and accurate evaluations [4]. Automated AI risk assessments can enhance the consistency, efficiency, and transparency of risk management processes, especially for high-risk applications like healthcare and finance [1]. As AI technologies continue to evolve, the automation of risk management becomes critical to safeguarding ethical and organizational standards and protecting stakeholders from unintended consequences.

To meet those challenges, we have developed an automated risk management technology to enable organizations to fully capture the underlying risk in their AI systems end-to-end. The overall workflow of our technology is visualized in Figure 1 and derived from the general risk management standard described in [2]. The three main pillars of the workflow are risk identification, followed by risk analysis or quantification, while the last step represents measurements and actions towards managing risks in AI systems. The whole workflow is designed as a circular process, involving a constant review. This review is not only limited to changes in potential risks but also to a changing landscape of regulation and standards which might require re-assessments and re-calculation of risks. Section 2 will give more details on the implementation of the described workflow. When implementing the full risk management workflow, certification or insurance of AI systems can be applied in a straight forward way. We represented those two aspects in dotted lines, as only AI systems under a specific risk category might require those aspects.

2 System Design

The design of our automated AI risk assessment tool incorporates a robust audit-proof mechanism, ensuring full traceability and accountability. Central to this system is a write-only architecture that guarantees the immutability of the audit trail, preventing any retroactive alterations or deletions. Every change made within the system is recorded in real-time, capturing essential details such as the nature of the modification, the identity of the user/role responsible, and precise timestamps. This comprehensive audit trail enables a clear reconstruction of the decision-making process and system evolution over time, ensuring transparency and compliance with regulatory standards. By maintaining a secure and tamper-resistant log, the system facilitates complete traceability, allowing auditors to verify compliance and transparency at any given point in time.

Additionally, the system features a comprehensive role model framework, designed to represent various stakeholders within an organization. This ensures that AI risks are assessed from diverse perspectives, including but not limited to technical, legal, and ethical viewpoints, aligning with internal organizational governance and regulatory requirements. The integration of this role-based approach enhances the depth and reliability of the risk assessment, ensuring that decisions are informed by a wide range of expertise and responsibility levels within the organization.

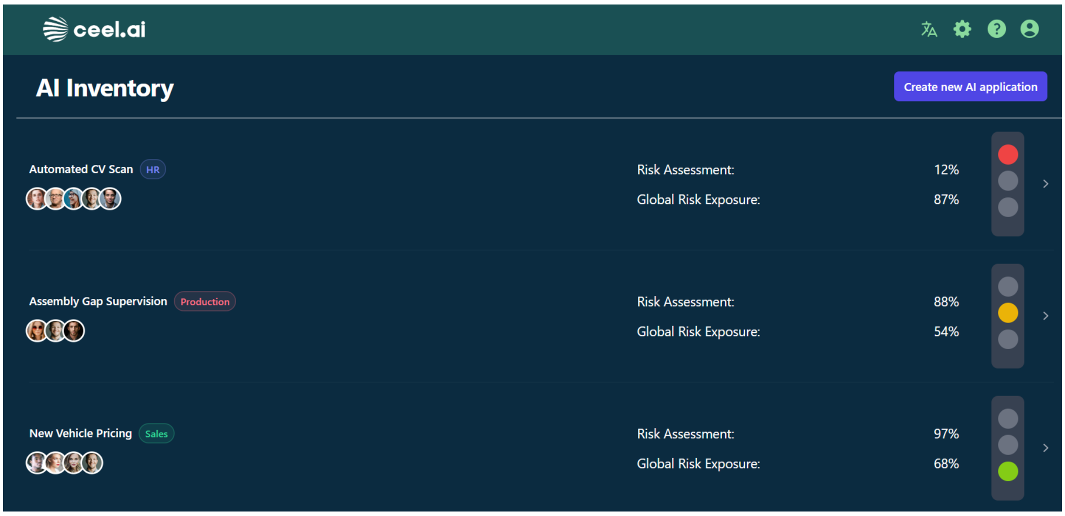

Figure 2 Inventory view in our software suite that collects all AI systems in an organization in one place.

The audit-proof and role based models are key design principles of our automation solution. Following Figure 1, we will describe two aspects of the AI risk management workflow in more detail, namely "Identify" and "Assess".

- **Identify:** The identification and collection of AI systems within an organization is a fundamental step in the successful assessment and risk quantification process. In our product suite, we provide an AI inventory, as illustrated in Figure 2. This inventory offers a comprehensive, high-level overview of all AI-based systems within the organization, centralized in one accessible location. Implementing an AI inventory enables organizations to track, monitor, and assess their AI assets effectively, which is crucial for managing risks associated with these technologies. Based on our experience working with companies of varying sizes, we have observed that the implementation process tends to be significantly more challenging for large corporations compared to SMEs or startups. This increased complexity arises from the larger number of departments and entities involved in the development of AI systems in large organizations, making coordination and oversight more difficult. As organizations continue to scale, maintaining an accurate and updated AI inventory becomes an essential component of risk governance.
- **Assess:** Each individual AI system collected in the AI inventory must be classified and ranked by risk category to ensure a structured and compliant risk management process. Figure 3 illustrates the assessment kernel implemented within our software solution, where the role-based framework plays a crucial role. Each AI system undergoes a comprehensive risk assessment to classify its risk management category in accordance with the requirements of the EU AI Act, while also providing a deeper understanding of its risk exposure. The software tool is designed to assess AI systems against both public standards, such as ISO and GDPR, as well as custom-defined standards when necessary to meet specific organizational needs. This flexible approach allows organizations to align the risk assessment process with their unique regulatory and operational requirements. As detailed in the system's overall architecture the assessment is audit-proof, ensuring full transparency and traceability for each AI system under review, thereby facilitating rigorous compliance and accountability across the entire AI lifecycle.

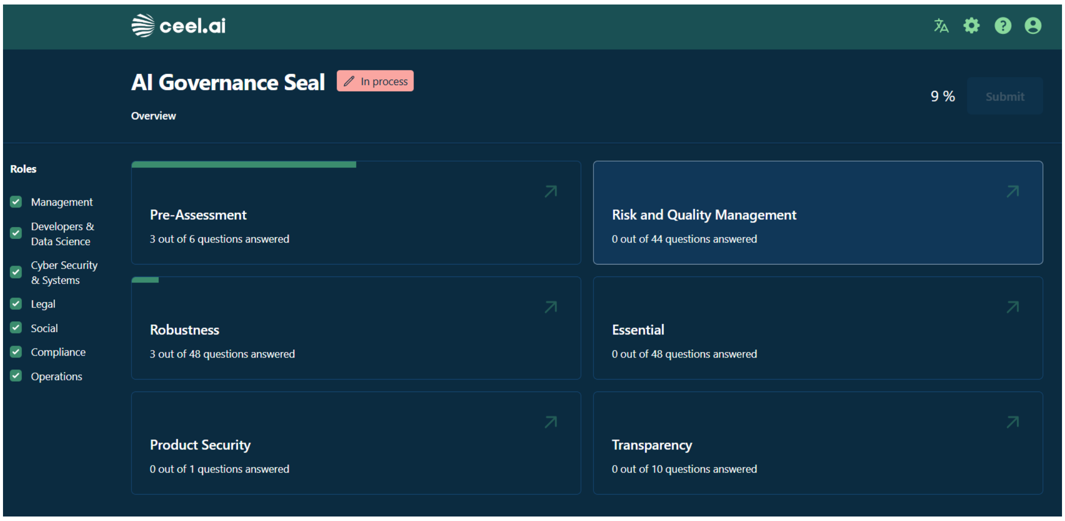

Figure 3 Assessment view, allowing assessments of AI solutions based on various standards and regulations.

3 Outlook

In this contribution, we have introduced a circular AI risk management workflow and an underlying software solution that automates AI assessments within an end-to-end framework. Central to this framework are traceability and multi-role setups, which, from a system architecture perspective, are essential to meet the requirements of existing and forthcoming regulations and standards. Moreover, we emphasize that risk analysis is not a one-time task but a continuous, circular process requiring the identification of new risks and the ongoing implementation and measurement of regulatory compliance.

Our experience working with organizations of varying sizes reveals significant uncertainty about how to address AI regulation and where to begin. To address this challenge, we recommend a straightforward approach:

- **Identify** all AI systems in your organization (Inventory).
- **Assess** the risk level of these systems in accordance with the EU AI Act.
- **Focus** on high-risk systems and ensure compliance with the relevant regulations.

We strongly believe that our framework, combined with our automation software solution, will simplify the compliance process for organizations striving to meet regulatory requirements. Looking ahead, we anticipate the development of additional standards and technical reports to provide detailed guidance for the successful assessment and certification of AI systems under the AI Act. Future research in AI risk management will prioritize enhancing processes for effective data collection in development and during operations of AI systems, ensuring that collected data are comprehensive, representative, and systematically gathered to support robust risk assessments. Additionally, emphasis will be placed on creating clear compliance frameworks to align with evolving regulations while promoting transparency and accountability. Finally, the establishment of continuous monitoring mechanisms will be crucial for enabling real-time risk detection and adaptive mitigation, ensuring that organizations can respond to the dynamic nature of AI systems and their associated risks.

References

1 R. Binns. Fairness in machine learning: Lessons from political philosophy. *CoRR*, abs/1712.03586, 2017. `arXiv:1712.03586`.

2 International Organization for Standardization. Iso 31000:2018 risk management. Technical report, ISO, 2018.

3 European Parliament and the Council of the EU. Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu) 2020/1828, 2024.

4 M. U. Scherer. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. *European Journal of Risk Regulation*, 29(2):354–400, 2016.