

Framework Summary

Establishing a Practical
Organizational Framework for
AI Accountability

Introduction

Artificial Intelligence (AI) has emerged as a transformative technology, promising advancements in various domains, from healthcare to finance and beyond. However, the rapidly growing interest in integrating AI systems into sometimes critical processes reveals a significant challenge: accountability. This policy brief presents a framework to help organizations establish and strengthen AI accountability practices. Our proposed organizational model draws from best practices and research to provide clear and practical guidance.

This brief shall serve as a basis for evaluating our framework and a mechanism for seeking input and feedback from industry, government, and academic stakeholders. We aim to advance real-world accountability efforts related to AI and understand how our work can further assist current initiatives. *Thus, we would greatly appreciate any [feedback](#) on our proposed solutions to evaluate their practicability.*

The Problem: Accountability Issues with AI

As a fundamental principle, accountability entails taking responsibility for actions and providing satisfactory justifications.¹ However, in the realm of AI, establishing accountability becomes complex due to the distinctive attributes of AI systems. AI systems operate on complex algorithms, vast datasets, and self-learning capabilities, making it often more difficult when compared to human decision-making to decipher the reasoning behind specific AI-generated outcomes. Given this lack of

interpretability and opacity in how AI systems derive their results and the prevailing view that AI cannot be held directly accountable for its actions, developers, implementers, and users of AI systems are left in a precarious position. This helps explain why the practical implementation of AI accountability still needs to be solved - the attributes that make AI powerful also challenge traditional notions of responsibility. Determining accountable human actors is, therefore, inevitable. This adds a layer of complication, as humans must account not only for their role but also for the behavior and impact of the resulting AI system.

In this context, the need for precise accountability mechanisms in the AI industry raises concerns about organizations' decision-making processes and the overall trustworthiness of the AI products they offer or implement. The importance of developing frameworks to navigate the complex, real-world challenges of ensuring accountability for AI systems is paramount. To understand the real-world problems of developers and deployers in integrating accountability, we investigated the critical issues at the core of the problem.

Through our research and discussions with practitioners, we found **three fundamental problems** that need to be solved for meaningful progress on accountability in the AI industry. First, there needs to be more **clarity and consensus** around what kind of accountability is required in specific operational contexts.² Indeed, while some initial frameworks for AI accountability have been proposed, there still needs to be standardized, cross-industry perspectives and guidelines that can be pragmatically implemented.³ This lack has led to a vicious

¹Cambridge Dictionary. (2022, May 11). Accountability. <https://dictionary.cambridge.org/dictionary/english/accountability?q=Accountability>

² Stix, C., (2021). Actionable principles for artificial intelligence policy: Three pathways. *Science and Engineering Ethics*. 27(15), 1-15.

³ Tekathen, M., and Dechow, N. (2013). Enterprise risk management and continuous re-alignment in the pursuit of accountability: a German case. *Manag. Account. Res.* 24, 100–121. doi: 10.1016/j.mar.2013.04.005; Pollmann, M. M., Potters, J., and Trautmann, S. T. (2014). Risk taking by agents: the role of ex-ante and ex-post accountability. *Econ. Lett.* 123, 387–390. doi: 10.1016/j.econlet.2014.04.004

Key Issues

1. The **lack of clarity around the practical implications** of AI ethics creates a vicious circle that hampers progress in AI accountability, leading to a shortage of best practices for implementing AI ethics, perpetuating the lack of clarity.
2. **Conflicts between the causes for the lack of exemplary AI accountability implementations**, such as limited understanding of what it entails, financial constraints, lack of incentives, and regulatory gaps, further complicate the resolution of the accountability issue.
3. The current absence of robust regulations and best practices has resulted in a **wait-and-see attitude** among many organizations, delaying proactive efforts to address accountability challenges.

cycle, where the scarcity of real-world case studies sustains the ambiguity. Second, the **underlying obstacles** preventing accountability initiatives - such as limited understanding of the concept in the AI field, budget constraints, misaligned incentives, and regulatory uncertainties - are complex and interrelated. Third, and perhaps most concerning, the current absence of standardized best practices and oversight has prompted a restrained response from many organizations. Without clear guidelines on compliance, some have adopted a cautious **wait-and-see stance** that postpones urgently needed work.

While some work is being carried out on the regulatory side, issues remain. For example, New York City established a task force to evaluate algorithmic systems.⁴ The

proposed EU AI Act⁵ guides risk assessments throughout a system's life cycle. These initiatives reflect a growing effort to unambiguously define accountable stakeholders and their duties to address theoretical obstacles and urgent real-world needs around AI governance.⁶ Regardless, our findings highlight that **examples of robust, yet adaptable accountability frameworks are still needed.**⁷

Overcoming these systemic issues requires unprecedented collaboration between stakeholders. Policymakers need to provide clear regulations and responsibilities to designers and providers. Jurisdictions have begun exploring what this accountability looks like in practice. Companies need practical tools to faithfully implement accountability measures now rather than waiting indefinitely. Academia can play a role by assessing accountability frameworks and facilitating knowledge-sharing. With a coordinated, multi-faceted approach, we can shift the tide and uphold AI development that respects the public interests of ethical and trustworthy AI.

Towards Improving Accountability

Our framework aims to promote trustworthy AI development, ensuring that providers and developers remain accountable for their actions while inspiring public trust and confidence in this transformative technology.

Our proposed organizational framework can directly address the **three main issues we identified concerning practical accountability implementation**. First, by providing precise and straightforward guidance, the clarity aspect of the framework aims to overcome the current

⁴ New York City (2018). Automated Decision Systems Task Force. Available at: <https://www1.nyc.gov/site/adstaskforce/index.page>.

⁵ Regulation (2021). Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (AI act). European Parliament, Council of the European Union. Available at: <https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX%3A52021PC0206>

⁶ Loi, M., & Spielkamp, M. (2021, July). Towards accountability in the use of artificial intelligence for public administrations. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 757-766).

⁷ Hohma, E., Boch, A., Trauth, R., & Lütge, C. (2023). Investigating accountability for Artificial Intelligence through risk governance: A workshop-based exploratory study. *Frontiers in Psychology*, 14, 1073686.

Key Solutions

- 1. Clarity:** The framework strives to provide straightforward guidelines for organizations developing and delivering AI systems, enhancing transparency in decision-making processes.
- 2. Accessibility:** By offering a low-effort-level approach, the framework will serve as a starting point for organizations to implement AI accountability effectively.
- 3. Comprehensiveness:** The framework will incorporate and summarise current efforts and state-of-the-art to foster accountability to ensure its comprehensive, addressing various aspects of the AI development decision-making process.

lack of definitional agreement that perpetuates ambiguity. Its **low-effort-accessible nature** responds to challenges like limited knowledge and understanding constraints by lowering barriers to adoption. Further, a **comprehensive yet cohesive synthesis of state-of-the-art research** helps tackle the fragmented landscape of partial solutions and inconsistencies between stakeholder interests that underlie accountability delays. Making it easier for organizations to promptly establish ethical best practices reduces the disincentives for proactivity

that stem from an uncoordinated regulatory environment. Finally, the framework's emphasis on synthesizing current resources aims to **boost standardization and coherence efforts**. This encourages moving past a restrained "wait-and-see" approach by supplying requisite regulatory and implementation support upfront. In sum, our solutions directly confront obstacles like conflicting understandings and lack of pragmatic guidance through an integrated framework design grounded in principles of clarity, accessibility, and comprehensiveness.

Our proposed accountability framework

Our framework proposes a solution to the definition of accountabilities by **embedding AI ethics requirements at each step of the AI development lifecycle**. Accountability is understood as taking responsibility and providing justification for one's actions. Therefore, it is implemented in a risk-based manner, identifying **risks for AI ethical principles** at each step and ensuring prevention or mitigation measures. The core idea is to **break ethical obligations down to more concrete actions** for which responsibilities and, thus, accountabilities can be defined more clearly. This process is accompanied by

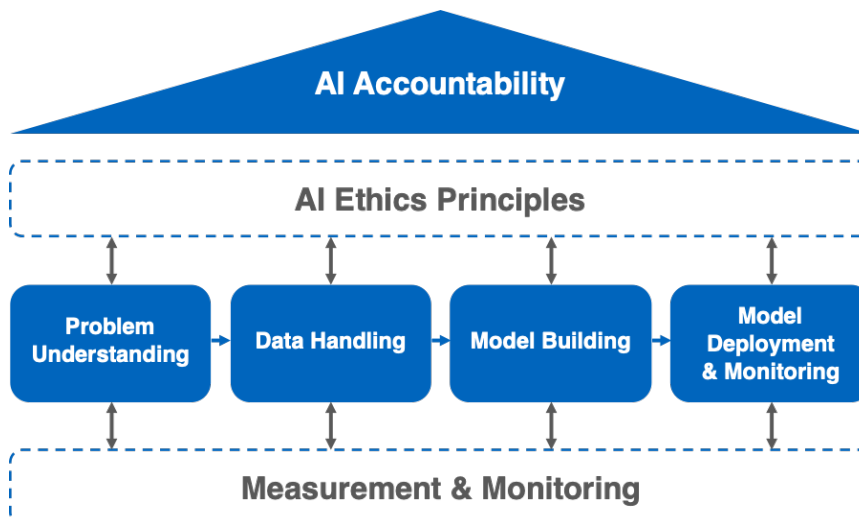


Figure 1: Accountability framework for AI systems.

ongoing measurement and monitoring to allow for faster reaction in case of potentially identified harm.

AI Accountability

Diving into the challenges of accountability for AI systems, we identified **three clear and effective accountability mechanisms** for developers and deployers. First, as regulations and ethical standards continue maturing globally, accountability is becoming more apparent - but more than simply achieving legal compliance is needed. **Fostering trustworthy and ethical practices** throughout an AI system's entire lifespan ensures it is developed and used in a way that respects societal norms. This thoughtful approach can proactively help mitigate potential harms. Secondly, **assigning responsibility** for AI outcomes is complicated by the many players involved at different stages. Developers, users, oversight bodies, and employing

organizations play interlinking roles. Figure 2 sketches the roles involved in AI development. Determining who should answer for what and how duties should be shared between stakeholders presents ongoing challenges that demand resolution, as discussed. Thirdly, **open communication** is vital for establishing trust and public acceptance of AI. Companies need to articulate their decision-making processes to users and communities clearly. They must also provide visibility into a system's true capabilities and limitations. By transparently demonstrating their dedication to responsibility, firms hope to contribute solutions that society is willing to utilize. Further, this transparency to the public facilitates the accountability mechanism globally in promoting individuals' and users' understanding of possible company misbehavior, but also rewards positive efforts. These three mechanisms, therefore, warrant immediate

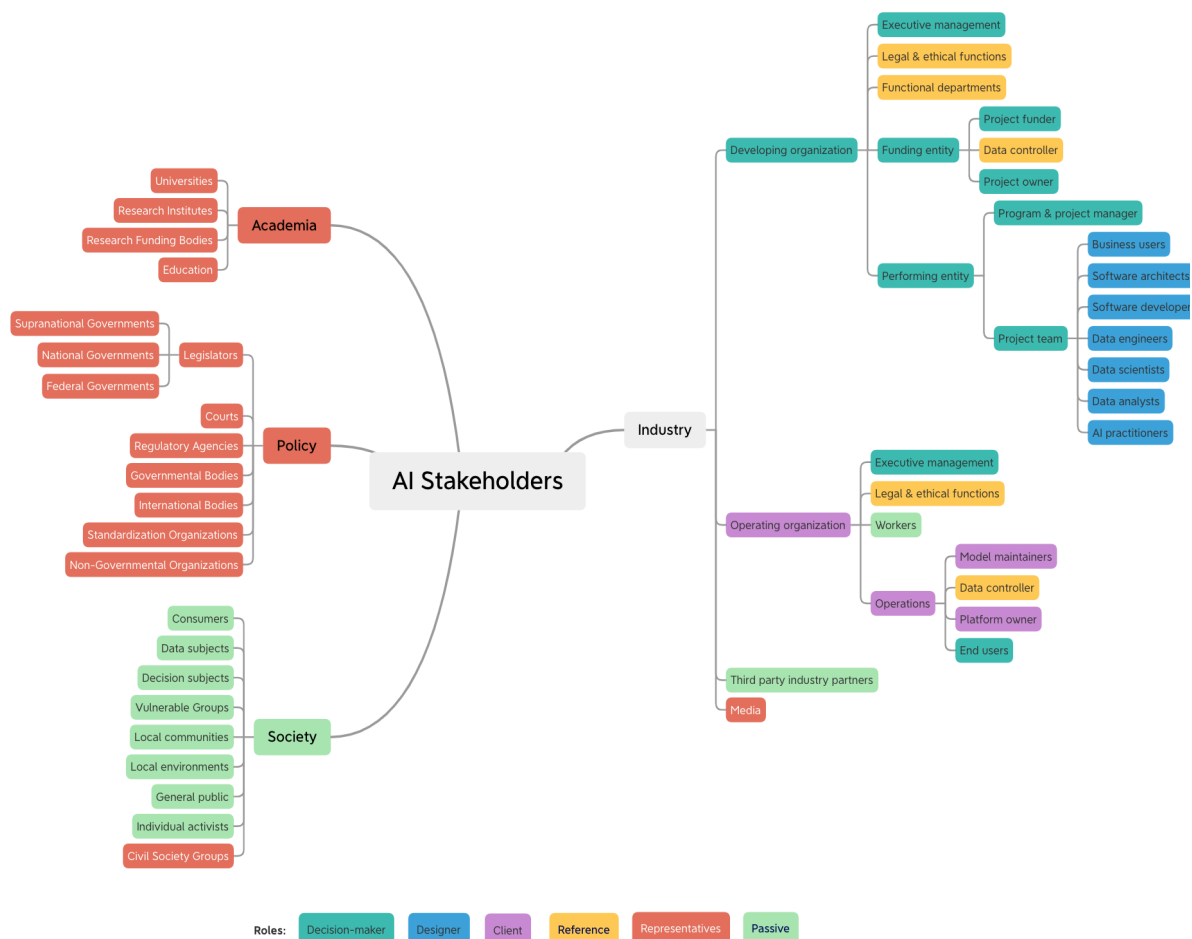


Figure 2: Map of stakeholders involved in the development and use of AI systems.

attention from those striving for accountability.

Our framework offers a **set of measures** to govern accountability across the entire AI lifecycle. Rooted in AI ethics principles, these measures are designed to be actively monitored and controlled. First, **processes across the AI lifecycle should be transparent**. We emphasize the importance of transparent development and deployment processes, where all responsible parties clearly outline their roles and tasks. By promoting transparency, companies can foster trust and ensure accountability is upheld at every step of the AI lifecycle. Moreover, clear and well-defined processes enable stakeholders to understand their responsibilities and contribute effectively to developing and deploying ethical AI systems. Second, practices of **stakeholder engagement** should be integrated. Accountability cannot be achieved in isolation. We highlight the critical role of stakeholder engagement in promoting accountability. This includes involving various stakeholders in the AI lifecycle, from developers to users. By assigning and communicating responsibilities, companies can build trust and foster a collaborative environment where all stakeholders actively participate in ensuring ethical and accountable AI practices. Engaging stakeholders throughout the process also allows for diverse perspectives and helps address potential biases or blind spots that may arise. Thirdly, **integrating ethics into AI development** is a core component of our framework. We emphasize the infusion of ethical principles into every stage of the AI lifecycle, ensuring alignment with societal values and norms. By incorporating ethical considerations from the outset, companies can proactively address potential ethical challenges and mitigate risks associated with AI systems. This integration includes robust ethical review processes, consideration of potential impacts on individuals and society, and adherence to

established ethical frameworks and guidelines. Finally, **continuous monitoring** of the technology. Accountability is not a one-time event but an ongoing process. We advocate for constantly monitoring AI systems to ensure adherence to ethical and regulatory standards throughout their lifecycle. This monitoring enables the timely identification of deviations from established norms and facilitates prompt adjustments or interventions to rectify ethical or legal concerns. Companies can continuously monitor AI systems to demonstrate their commitment to accountability, minimize potential harm, and maintain public trust.

By systematically addressing the challenges outlined in our findings and implementing the concrete solutions provided, organizations can create technologically advanced AI systems that are also ethically accountable. Our framework aligns with the principles established by international organizations, regulatory bodies, and AI policies, empowering companies to navigate the intricate landscape of AI accountability with confidence and responsibility.

AI Ethics Principles

In our pursuit of establishing a comprehensive and practical framework for AI accountability, we approach accountability as a set of responsibilities that result from the need to adhere to the ethical principles of AI. These responsibilities or obligations are addressed through a **risk management approach**, which lays the foundation for later defining stakeholders' roles and responsibilities in risk prevention and resolution processes.

Existing risk-based initiatives towards ethical AI and related problems

Our approach draws inspiration from existing international and business-level risk management frameworks designed for AI systems. Governments and international organizations have been actively working

Key mechanisms to enable accountability assurance

1. Transparent Processes

We emphasise the importance of transparent development and deployment processes, where the responsible parties clearly outline their roles and tasks.

2. Stakeholder Engagement

We highlight the involvement of various stakeholders, from developers to users, in promoting accountability. Assigning and communicating responsibilities is integral to building trust.

3. Ethics Integration

Our framework guides the infusion of ethical principles into every stage of AI development, ensuring alignment with societal values and norms.

4. Continuous Monitoring

Accountability is an ongoing process. We thus advocate for continuous monitoring of AI systems, enabling prompt adjustments in case of deviations from ethical or regulatory standards.

to create risk management frameworks for AI systems, aiming to outline best practices and provide clarity, such as the considerations on requirements for trustworthy AI by the European Union's High-Level Expert Group on AI (AI HLEG)⁸, the OECD⁹ classification framework for AI systems, or, as a result, the European Commission's proposal for a risk-based classification of AI technologies in the AI Act proposal¹⁰. These frameworks acknowledge the need for tools to mitigate risks. However, they fall short of

accommodating the diverse range of AI systems, impeding the creation of tailored solutions for different AI technologies.

While **no universally applicable risk management framework exists** for all AI systems internationally, there is a consensus on specific criteria for AI risk identification regardless of the framework used. These criteria include evaluating the potential risks' scale, scope, and optionality.¹¹

However, in the interest of developing an adequate solution for practitioners, several **critical gaps in the current risk-based approach** to accountability for AI systems have been identified.

First, the mutual effects of **insufficiently defined accountability and the lack of clearly distributed risk management measures** hinder the proper definition of both. The definition and assignment of accountabilities are pivotal in shaping risk management measures. The ambiguity surrounding responsibility for risk hinders the effective determination of who should be accountable for its mitigation. With clear lines of responsibility, it becomes easier to implement appropriate risk management strategies and allocate necessary resources. Resolving this ambiguity is crucial to establishing a robust framework that ensures accountability throughout the AI system lifecycle.

Second, **transparency** emerges as a significant challenge in the governance of AI risks. The inherent complexity of AI systems poses difficulties in determining appropriate response mechanisms to identified risks. Additionally, unclear

⁸High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019). Ethics Guidelines for Trustworthy AI. Brussels: European Commission. Available at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

⁹OECD. (2021). Tools for trustworthy AI : A framework to compare implementation tools for trustworthy AI systems. Documents de travail de l'OCDE sur l'économie numérique, n° 312, Éditions OCDE, Paris, <https://doi.org/10.1787/008232ec-en>

¹⁰ Regulation (2021). Regulation of the European Parliament and of the Council Laying Down

Harmonized Rules on Artificial Intelligence (AI act). European Parliament, Council of the European Union. Available at: <https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX%3A52021PC0206>

¹¹ OECD. (2022, February). OECD Framework for the Classification of AI Systems. OECD Digital Economy Paper. <https://www.oecd-ilibrary.org/docserver/cb6d9ecaen.pdf?expires=1652269451&id=id&accname=guest&checksum=CFE5F318317EA246D4F21C90EB75BEDC>

governance measures further impede the effectiveness of transparency initiatives. This lack of clarity can lead to a lack of trust and understanding among stakeholders, hindering the adoption of risk mitigation strategies. Furthermore, the need for more awareness and expertise among operators and practitioners exacerbates the transparency challenge. Addressing these transparency challenges requires comprehensive guidelines and accessible resources to enhance understanding and promote transparency throughout the AI ecosystem.

Thirdly, **managing unanticipated events and consequences** is critical to AI accountability. These unforeseen risks demand focused attention and targeted mitigation efforts. The dynamic nature of AI systems and their interactions with complex real-world environments make predicting and preparing for all potential consequences challenging. To address this challenge, it is essential to establish mechanisms for ongoing monitoring, proactive risk assessment, and continuous adaptation. This includes the development of robust incident response protocols, clear escalation procedures, and cultivating a

culture of learning from unanticipated events.

In sum, we can shape a more effective and comprehensive solution by recognizing these critical deficiencies in the current risk-based approach to AI accountability.

Requirements for effective risk governance frameworks

Drawing on the insights gathered from our research, we have discerned **five key requirements** that should inform AI risk governance approaches.

First and foremost, achieving **balance** is of paramount importance. We recognize the need to strike a delicate equilibrium between specialized and generalized risk management processes. Our approach aims to be adaptable across sectors while acknowledging the unique organizational contexts in which AI systems operate. By striking this balance, we can establish a comprehensive framework that addresses sector-specific risks while leveraging common risk management principles.

Second, the **extendibility** of risk management approaches is crucial. Given the rapid evolution of the AI landscape, our

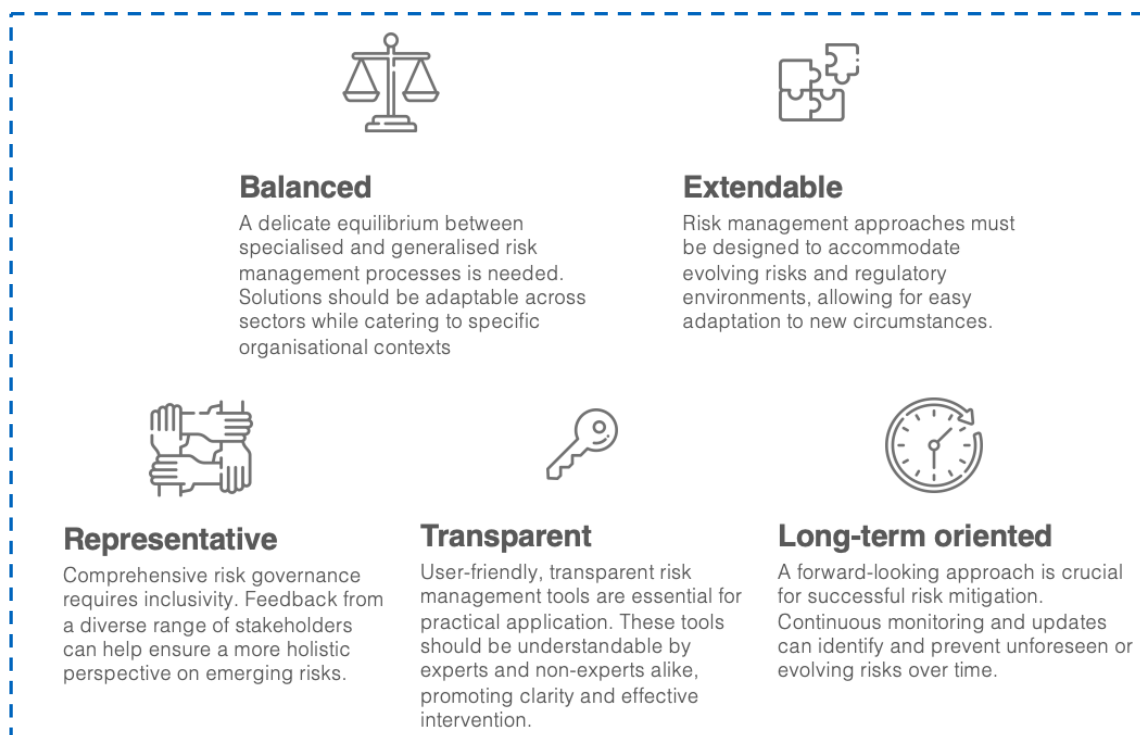


Figure 3: Requirements for Risk Governance Frameworks.

framework is designed to accommodate evolving risks and regulatory environments. Organizations can proactively address emerging risks and remain compliant with changing regulations by ensuring adaptability to new circumstances. Comprehensive risk governance necessitates representation.

Third, we emphasize the importance of soliciting **representative feedback from diverse stakeholders**, including experts, practitioners, users, and affected communities. Incorporating these diverse perspectives enables us to gain a more holistic understanding of emerging risks and develop risk management strategies that serve the interests of all stakeholders involved.

Fourth, **transparency** is a cornerstone of effective risk governance. Our framework prioritizes the development of user-friendly and transparent risk management tools. These tools are designed to be easily understandable by experts and non-experts, promoting clarity and facilitating effective intervention. By fostering transparency, we enable stakeholders to actively participate in risk governance and decision-making processes.

Finally, a **long-term orientation** is essential for successful risk mitigation. Our approach emphasizes continuous monitoring and updates to identify and prevent unforeseen or evolving risks over time. By adopting a forward-looking perspective, organizations can proactively address emerging challenges and ensure the ongoing trustworthiness of their AI systems.

In conclusion, as the landscape of AI technologies evolves, the need for practical methodologies and frameworks for responsible AI becomes increasingly pronounced. Our research, transitioning from theoretical definitions to actionable methodologies, aligns with this demand. We've identified pressing questions about AI accountability and suggested a risk-based approach to address them. This

approach leverages the interplay between risks and responsibilities in line with established risk management methodologies, regulations, and frameworks.

A Trustworthy Development Process for AI Systems

While defining accountabilities, i.e., concrete responsibilities and the ability to justify related actions, is difficult at the level of AI ethics principles, it becomes more feasible when the principles are broken down into concrete steps for risk mitigation. Therefore, this framework's heart is a **development process describing the actions required for responsible or trustworthy AI systems**. An emerging consensus has been found in conceptual and practice-oriented literature around the measures that can be implemented during system development to prevent or reduce risks during system operation.

Measures are required at two levels. Activities related to strategic decision-making and giving guidance on the ethical development of AI systems from a general perspective must be defined broadly on an **organizational level**. Such measures include, e.g., the development of an organizational AI governance strategy or the creation of codes of conduct. Further, if internal or external validation of ethical behavior is required, a strategy or entity can be defined for the organization as a whole. Guidance on supporting education on AI practices inside or outside an organization can be laid down independent of concrete projects.

In contrast, specific risk management and mitigation measures highly depend on the particular use case or context and, therefore, are defined on a project level. These measures can be structured along the AI development lifecycle and arise from four activity categories: (1) planning, (2) assessment & ensuring, (3) creation, and (4) communication.

On a project level, **planning activities** are required to define project objectives and align them with the strategic company or project goals. For example, they include measures to specify responsibilities or determine system requirements or thresholds. Mechanisms for **assessing compliance** with certain obligations of AI providers and taking appropriate action where needed include, for example, standardized pre- and post-development impact assessments (incl. technical testing) on specific system properties and impacts, as well as assessment of system dependencies, the legality of data processing and development, team properties and capabilities, options for audits (incl. seals of approval and certifications), truthfulness and ensuring remuneration. **Creation activities** relate to obligations that require or impact the system's active (re-)design or organizational processes, such as participatory development or public intervention mechanisms. Lastly, **communication activities** are linked to the disclosure of certain information, communication of system definitions, purpose, limitations, risks, and use, as well as education of staff, users, and the general public.

Such a process for the trustworthy development of AI systems outlines the measures required to mitigate and manage risks and, therefore, clarify the responsibilities that arise with the endeavor to build trustworthy AI. Transparently following and communicating such a process can help justify measures and thus significantly contribute to the definition of accountabilities. Finally, they also indicate the responsible actors or roles, depending on their stage in the development life cycle and the related development activities with which they are associated.

Measuring & Monitoring

To ensure proper realization and true impact, measures to increase accountability should be accompanied by **ongoing measurement and monitoring**. Here, it is particularly important to consider how the measures impact the ultimate goal of embedding ethical principles in the AI system. Therefore, instead of monitoring the individual actions taken, their overall effect can be observed by continuously measuring the resulting system's ethicality by assessing their adherence to ethical principles. To do so, ethical principles must

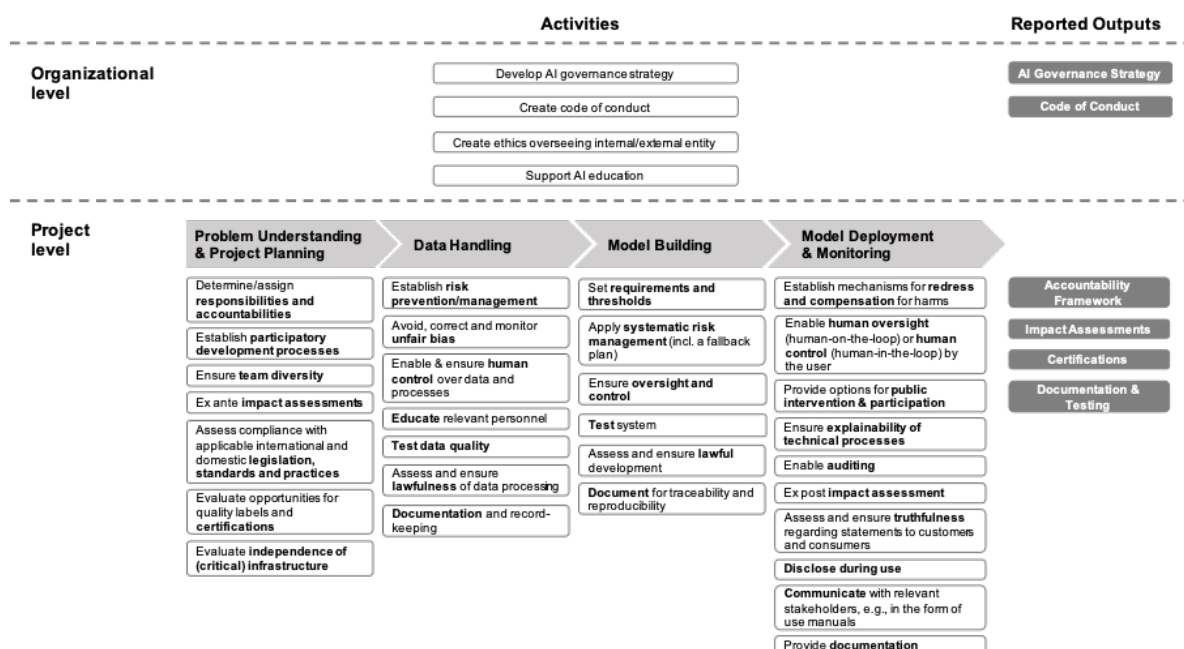
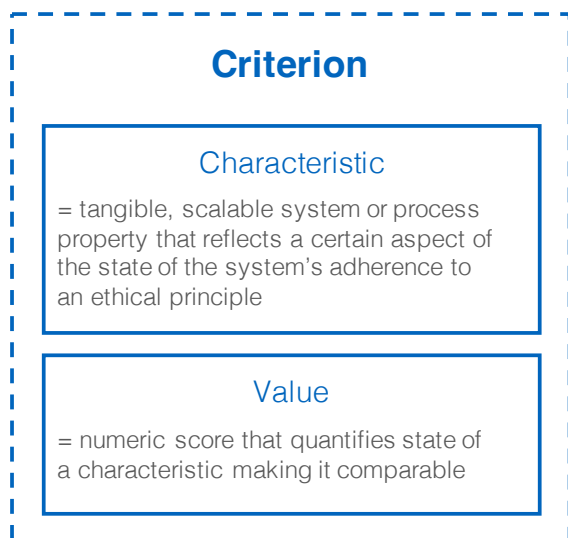


Figure 4: A trustworthy development process for AI systems.

be quantified to be measurable and scalable to observe their alteration.

In our framework, we propose a **method to quantify and hence be able to measure variations in the realization of ethical principles** in the definition of criteria consisting of tangible, scalable characteristics and an associated target value. **Characteristics** are those system or related process properties that reflect the current state of a system's adherence to a given ethical principle. For example, when measuring an AI system's transparency, the ratio of user inquiries out of all user requests related to issues of understandability could be determined as a suitable characteristic. They should be tangibly scalable to allow the assignment of a numeric value. These **values** are the scores that relate to an identified characteristic and quantify the AI system's status regarding ethical principle compliance. A scale to measure an optimal state shall be defined and assigned to each characteristic.

For example, in the case of measuring an AI system's state regarding transparency by assessing the ratio of user inquiries that relate to understandability, a value of only 10% might be considered acceptable. In comparison, a value of 60% could reflect a



¹² UNESCO. (2021, November). Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

¹³ High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019). Ethics Guidelines for Trustworthy

critical threshold indicating that the system is not regarded as transparent enough. Continuously comparing the determined relevant system properties against the developed scale allows us to measure compliance with the targeted ethical principle implementation.

Both characteristics and values must be determined based on the specific use case, system, and context to identify relevant system requirements and thresholds. It is, therefore, during model building at the latest that such a measurement should be initiated. Regular comparison and alignment before and during development as well as during deployment, use, or reiteration of the system allows monitoring of the system's ability to adhere to the defined ethical principles and indicates red flags, thus, calls for action.

Conclusion

The difficulty of defining accountabilities for the outcomes of AI systems has frequently been highlighted as a severe one. The lack of clarity around actions for AI providers required for solving this problem, together with the lack of best practices and therefore an often-observed wait-and-see attitude, further hinders an effective resolution. Thus, there has been a call for solutions to enhance **clarity**, **accessibility**, and **comprehensiveness** around required measures to support accountability.

However, determining accountability in practice remains an open challenge, as acknowledged by international bodies like UNESCO and the EU's High-Level Expert Group on AI.^{12,13} As the EU recognizes, ensuring AI systems are fair, aligned with ethical values, and have suitable governance is vital as these technologies

AI. Brussels: European Commission. Available at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

increasingly impact many lives.¹⁴ The OECD definition further specifies that **accountability requires due attention throughout an AI system's lifecycle by all relevant actors.**¹⁵

To account for these challenges, regulatory efforts are undertaken worldwide (e.g., in the EU¹⁶, Brazil¹⁷, or Japan¹⁸). While legal frameworks are inevitable in the discussion to ultimately clarify and unify the current diversity of AI governance streams, their provisions can always only guide practical implementations. It is the legislator's responsibility to set the clear underlying principles and potential red flags, however, naturally, they lack in low-level accessibility. For example, refinements have been demanded on the workability of the EU AI Act proposal.¹⁹

Detailed policy recommendations and standards are a reliable source of more detailed guidance for the regulatory provisions, and they are currently under development for ethically aligned AI (e.g., in the EU²⁰ or Australia²¹). Standards, like the IEEE's Ethically Aligned Design framework²² detail the requirements to reach responsible AI governance and present accountability as an integral part. However, largely being under development, they often do not offer clear step-by-

step guidance but general recommendations to reach ethical and accountable AI design, development, or deployment.

By founding our framework on ethics principles applied at each stage of development and use, we aim to operationalize our conceptualization of accountability as a **holistic and integrated responsibility**. Taking a lifecycle view that systematically enforces ethical norms from ideation through all subsequent phases, our framework aims to make accountability an integrated part of standard organizational processes for trustworthy AI. Therefore, the framework is **adaptable to the use case** of the developer and provider, as it comprehensively lays out the steps to take. Finally, this lifecycle approach involves implementing ethics best practices from the initial design stage through routine use. A vital component is establishing mechanisms to measure²³ and monitor adherence to these principles at each lifecycle stage, from ideation to post-market implementation.

In summary, our framework can support solving the issue of defining accountability for an AI system's actions. It fosters clarity by offering an overview of required measures and at the same time, these more concrete measures can define

¹⁴ European Council. (2020, August). Special meeting of the European Council – Conclusions (EUCO 13/20).

<https://www.consilium.europa.eu/media/45910/021020-euco-final-conclusions.pdf>

¹⁵OECD. (2021). Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems. Documents de travail de l'OCDE sur l'économie numérique, n° 312, Éditions OCDE, Paris, <https://doi.org/10.1787/008232ec-en>.

¹⁶ Regulation (2021). Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (AI act). European Parliament, Council of the European Union. Available at: <https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX%3A52021PC0206>

¹⁷ De Agência Senado (2022). Comissão de juristas aprova texto com regras para inteligência artificial. Senado Federal. <https://www12.senado.leg.br/noticias/materias/2022/12/01/comissao-de-juristas-aprova-texto-com-regras-para-inteligencia-artificial>

¹⁸ Japan METI (2021). 'Governance Guidelines for AI Principles in Practice Ver. 1.1' compiled (METI/Ministry of Economy, Trade and Industry).

Available at: <https://www.meti.go.jp/press/2021/01/20220125001/20220124003.html>

¹⁹ German AI Association. (2023). Towards the finish line: Key issues and proposals for the trilogue .. German AI Association. https://ki-verband.de/wp-content/uploads/2023/07/Position-Paper_AI-Act-Trilogue_GermanAIAssociation.pdf

²⁰ European Commission High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

²¹ Australian Government. (2019). Australia's Ethics Framework. A Discussion Paper. Department of Industry, Innovation and Science. <https://www.csiro.au/en/research/technology-space/ai/ai-ethics-framework>

²² IEEE. (2019). Ethically aligned design. IEEE. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

²³ Mantelero, A. (2020). Elaboration of the feasibility study. Council of Europe.

accountabilities, as concrete tasks can be attributed to roles in the system development. The modular structure and the step-by-step breakdown of actions within our framework facilitate accessibility and practical implementation. In addition, the comprehensive nature of the framework enables system-independent adaptability, which makes the framework versatile and allows for broader adaptation to various use cases.

For more information on the framework development and methodologies, we refer to our [project webpage](#).

We would greatly appreciate your feedback and, therefore, kindly ask you to participate in our [consultation](#).