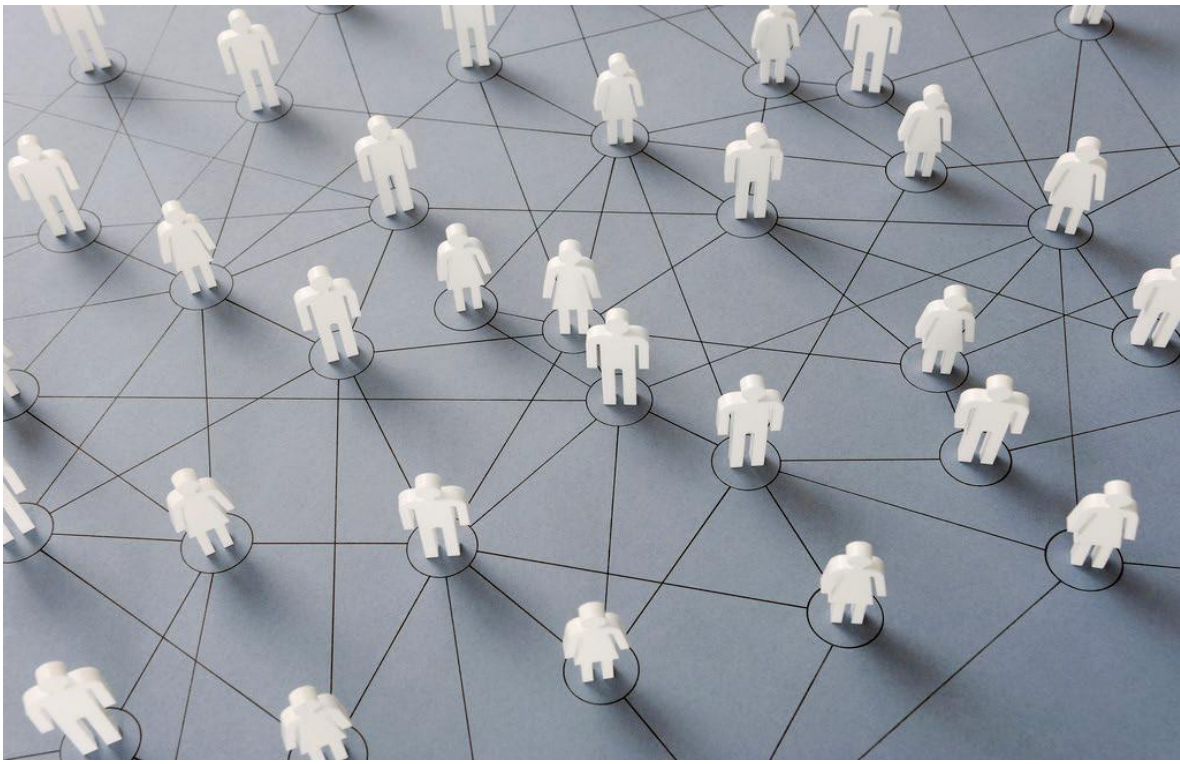
VU Risk & Crisis Management
Knowledge hub

School of Business and Economics for Professionals

AI governance

December 2024

10 key points in discussion with AI. An almost-scientific experiment.



INSIDE

—

10 key governance components

We provide commentary on the components as suggested by AI

—

New research in the spotlight

We highlight some of the recent papers on the topic

Authors

Julia Vitte
Director, risk management expert, conducting her research at VU Amsterdam.

j.vitte@vu.nl

Jost Sieweke
Director, Associate Professor , Programme director of the Executive MBA Leading with Purpose at VU Amsterdam.
j.sieweke@vu.nl

Executive Summary

Organizations in various sectors are moving towards embedding AI models of different types and complexity in their day-to-day operations. In this paper, we provide a high-level overview of governance elements, which are needed for effective and efficient AI usage, considering internal and external organizational stakeholders. Given the broad nature of AI governance, this paper introduces the core elements. In future white papers, we will zoom in each element .

We suggest that while the topic is broad, multiple elements are already addressed and developed in the organizations. The practice of assessing risk is common, as is the practice of regular audits. Other topics, for example mitigation of bias and discrimination in decision making or governing synthetic data in addition to "normal" data, are new and require attention and education.

We hope that the overview of governance elements introduced in this and future white papers helps risk managers and decision-makers to effectively implement AI, without losing control or implementing harmful practices.

KEY DEFINITIONS

*An **AI system** is a machine-based system that generates from its input (e.g., text, numerical data) outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.*

***Trustworthy AI [DO WE NEEED TO DEFINE TRUSTWORTHY AI FIRST?]** has three components: it should be (1) lawful, ensuring compliance with all applicable laws and regulations; (2) ethical, demonstrating respect for, and ensure adherence to, ethical principles and values; and (3) robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm. Characteristics of Trustworthy AI systems include: validity and reliability, safety, security and resilience, accountability and transparency, explainability and interpretability, privacy-enhancement, and fairness with harmful bias managed. Trustworthy AI concerns not only the trustworthiness of the AI system itself but also comprises the trustworthiness of all processes and actors that are part of the AI system's life cycle. Trustworthy AI is based on respect for human rights and democratic values.*

AI-generated image representing "governance for AI"

What we've asked Chat GPT:

Could you please tell me how would a company best govern their use of AI?

How would you visually represent AI governance in a company?

Setting up risk management frameworks for the newly emerging technologies is one of the key tasks of firms and risk managers. How to start with this task? We thought why not ask AI what it would recommend ensuring a robust and consistent AI governance framework – from managerial oversight to practical solutions, for example risk assessments or risk reports.

The picture above provides a visual representation of AI governance, according to AI. It is an interesting view, which we will explore further also in future white papers, but for now we are very happy about the emphasis on ethics!

AI answered our question regarding AI governance with a list of core points which need to be set up to effectively govern AI use in a company.

Below we aim to systematize the 10 key governance elements and assess each point and provide some comments.

AI suggests	***We comment***
1. Define Clear Objectives • **Vision and Mission**: Align AI projects with the company's mission, values, and business goals. • **Prioritization**: Prioritize areas where AI can add the most value while considering potential risks.	We agree with the first step of alignment of AI initiatives with the strategic goals. What we would like to explore further is how various strategic goals, which are formulated on a high level, would be aligned with the activities where AI is used, as those activities are usually executed on a rather operational level. An example is a company aiming to expand to a new market. At a strategic level, the goal is clear. Yet, the question is how to set up the right processes for AI to be implemented effectively and correctly at the operational level.
2. Establish an AI Governance Framework • **AI Ethics Policy**: Develop guidelines that outline ethical principles for AI use (e.g., fairness, transparency, privacy, accountability). • **Governance Committee**: Form a multidisciplinary team to oversee AI initiatives. Include members from IT, legal, compliance, data science, HR, and business units. • **Lifecycle Oversight**: Ensure governance covers the entire AI lifecycle—development, deployment, monitoring, and retirement.	This is a great point. A framework with clear roles and responsibilities will be crucial for AI governance. The complexity of such a framework will vary depending on company complexity. Our main advice here would be to re-use the existing governance framework for governing of AI. To do this, we need to evaluate the areas in which AI is used, and how those are governed already, and build up from there. An element which we would suggest adding is an oversight committee, which will critically check AI governance and make sure that company leaders are informed about development, implementation and use of AI models. An AI system lifecycle[1] typically involve several phases: (i) planning and design, (ii) data collection and processing, (iii) model building and/or adapting existing models to specific tasks, (iv) testing, evaluation, verification and validation, (v) deployment, and (vi) operation and monitoring. These phases often take place in an iterative manner and are not necessarily sequential, and require various types of oversight, which we will cover in the upcoming white papers.

[1] https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence-second-edition

3. Ensure Legal and Regulatory Compliance

- **Regulations**: Stay updated on AI-related regulations (e.g., GDPR for data privacy, industry-specific laws).
- **Periodic control**: Conduct regular audits to ensure compliance with laws governing data use, discrimination, and intellectual property.

Professionals who work with regulations know all too well how quickly the landscape changes, and that our supervisors and regulatory bodies keep issuing new and improved standards regularly.

The most recent "AI Act[2]", besides already known NIST and other regulations, sets up a legal framework for the development, the placing on the market, the putting into service and the use of AI systems. The regulation recognized that AI, while being overall beneficial, poses certain risks.

4. Data Governance

- **Data Quality**: Establish processes to ensure the data used for AI is accurate, complete, and unbiased.
- **Privacy Protections**: Implement measures like data anonymization and encryption.
- **Access Control**: Restrict who can access sensitive data and AI models.

Looking at the financial services industry, there is already quite a robust framework of policies, procedures, controls, and checks. Other industries are not far behind. While the industry is fairly used to governing data used for building mathematical models, AI introduces new types of data, which may require additional mechanisms, such as training data and synthetic data. Still, each company needs to ensure that core elements such as *segregation of duties, proper record retention, ensuring continuity, and process of data storage* are built into these new AI-related processes. This will require additional training for certain functions (e.g., risk managers need to understand more about data governance) if we look at larger companies, and the need to establish additional resources for smaller ones. So, while AI as such will optimize resources, governing AI may require additional ones.

[2] Official regulation name: REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

5. Focus on Ethical AI

- **Bias Mitigation**: Regularly test AI systems for bias and take corrective measures.
- **Explainability**: Ensure AI decisions can be explained to stakeholders, especially for high-stakes applications.
- **Human Oversight**: Require human review for critical decisions, such as hiring, lending, or medical diagnoses.

Ethics is crucial when it comes to AI usage. Of course, bias is a well-known term. Still, for clarity, we would like to scope the definition of bias: *An AI system's differential treatment of different groups, which may arise from implicit systems of meaning, norms and values*[3]. Bias mitigation may pose a practical challenge, as different types of harmful bias emerge due to a variety of factors, including but not limited to human or system decisions and processes across the AI lifecycle; preexisting cultural and social bias in training data; technical limitations (such as non-representative or limited design specifications and data); or use in unanticipated contexts*.

Confabulation is an interesting phenomenon of AI generating false responses, which may look plausible to the user.

The ability to control for these and many other factors will be key for successful use of AI.

6. Invest in Training and Education

- **Training**: Train employees to understand AI capabilities, risks, and limitations. Provide specific training for those managing and developing AI systems to adhere to governance policies.

To comment on this one – we agree.

From new terminology to new process logic, professionals across various fields are facing a field of knowledge previously reserved for developers and IT professionals. The educational landscape is already changing in response to the changing business needs. Yet, companies must invest in training their employees to increase the effectiveness of their AI usage.

7. Risk Management

- **Assess risks**: Conduct risk assessments before deploying AI systems to identify potential impacts on stakeholders.
- **Planning**: Develop contingency plans for AI failures or unintended outcomes.

Being a risk research center, we would like to pay special attention to this component. Beyond the current overall risk management methodology, or model risk management, the European Commission offers a new framework for self-assessment of AI trustworthiness, which should be leveraged. It requires human oversight, technical robustness and safety, requirements on privacy and data governance, transparency, diversity, non-discrimination and fairness as well as considerations for societal and environmental well-being. Finally, one of the most important aspects is accountability.

[3] https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence-second-edition

In the coming whitepapers, we will dive deeper into each of these components and offer practical ideas on applying the existing risk management methodologies to these requirements.

8. Implement Robust Monitoring

- **Measure**: Use performance metrics to track AI systems' outcomes and effectiveness.
- **Monitor**: Set up alert systems for anomalies or unexpected behavior in AI models

While the suggested points are fair, we see the need for having processes or built-in steps for AI system's auditability, from the traceability of AI development to documentation of certain choices/ assumptions, and further through the cycle, to assessment of AI outcomes and their impact.

9. Engage Stakeholders

- **Clear communication**: Communicate AI initiatives and their impact to stakeholders (e.g., employees, customers, and partners).
- **Stakeholder management**: Establish feedback channels for stakeholders to report concerns or suggest improvements.

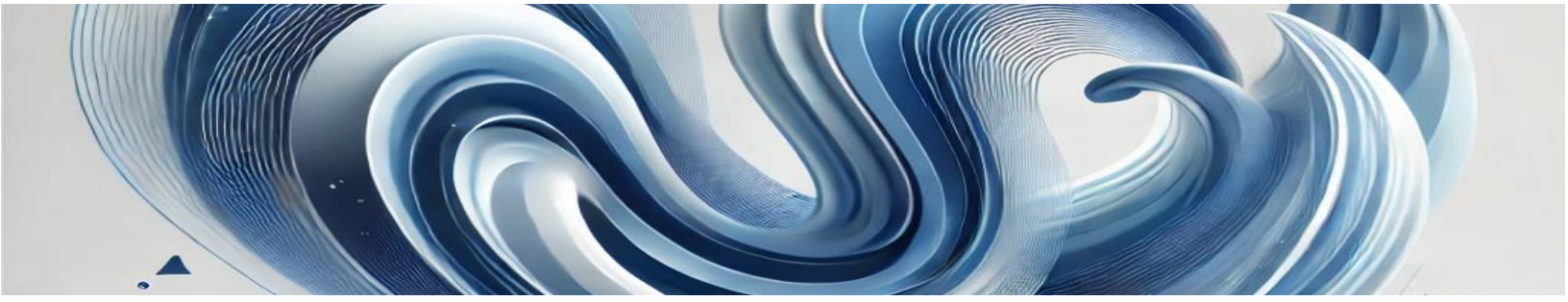
When it comes to AI, there is virtually no group of people who are *not* stakeholders. Internally – employees can be developers, users of AI models, or various control and support functions. Externally, we are talking about clients, but also about government bodies, supervisors, and auditors. Establishing feedback channels for these stakeholders is vital for companies to adapt their AI systems, if necessary, and to benefit from the experiences and learnings of their stakeholders.

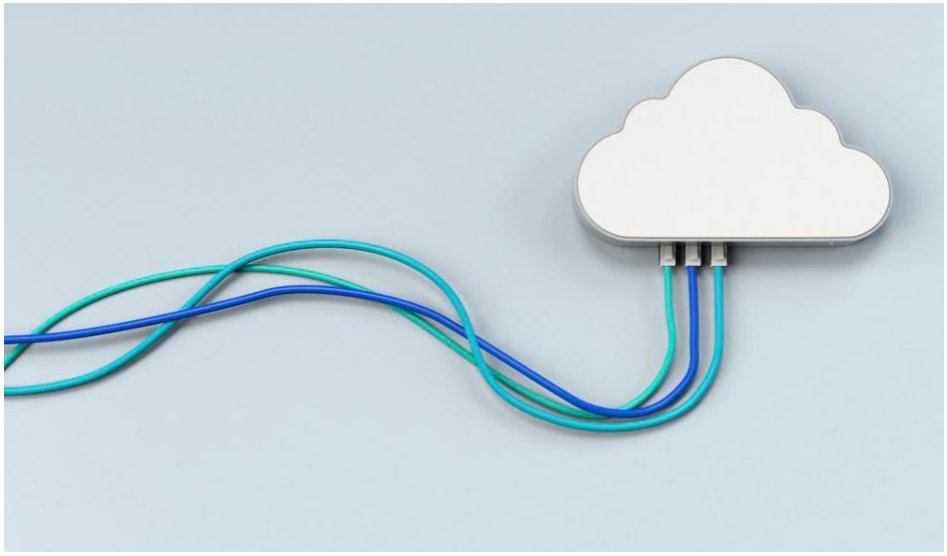
10. Audit and Continuous Improvement

- **Regularly review and update:** AI governance policies need to be up to date based on technological advancements, new regulations, and stakeholder feedback.
- **Evaluation**: Perform periodic audits to evaluate adherence to governance frameworks.

The official definition of auditability refers to the ability of an AI system to undergo assessment of its algorithms, data and design processes, in particular to determine whether the system is working as intended.

Too many times, we are faced with a legacy system that was never built for auditability. This is a great opportunity to build processes which surround AI, with auditability in mind.





In summary

As AI continues to impact various sectors, effective governance is essential to balance innovation with risk, promote ethical considerations, and ensure societal well-being.

While AI forms and tools are being used for various purposes in a multitude of organizations, it is important to recognize existing and future governance challenges. Organizations need to ensure that strategic vision is operationalized; that processes surrounding AI development are established with auditability in mind; and that control frameworks will be of the highest quality – sufficiently heavy to control for the risks but not too heavy that they hamper innovation and development speed.

What do we know from research?

Recent research on AI governance highlights the need for effective oversight to address ethical concerns and ensure responsible development.

To enhance trust and mitigate risks, organizations are developing governance frameworks which aim to ensure fairness, explainability, and accountability in AI deployment (Usmani et al., 2022), while promoting data stewardship, transparency, and risk-based controls (Janssen et al., 2020).

Current global efforts involve various stakeholders, including the private sector, public sector, and international organizations (Butcher & Beridze, 2019).

Future research agenda includes prioritizing AI safety research (Zhang et al., 2021).

About us

VU Risk and Crisis Management knowledge hub is a dedicated space for industry practitioners and academics to connect with each other. We have created it in response to growing global uncertainty, as well a disconnect we noticed between industry practices and latest research.

You can find more on our website and LinkedIn.

Julia Vitte

Director, risk management expert, conducting her research at VU Amsterdam.

j.vitte@vu.nl

Jost Sieweke

Director, Associate Professor , Programme director of the Executive MBA Leading with Purpose at VU Amsterdam.

j.sieweke@vu.nl

Julia and Jost love a good conversation!
Feel free to approach us, even if we have never met.