

Research Topic: Developing Governance Frameworks to enhance trust in AI Tools across large organizations.

As AI continues to transform industries, it is increasingly important to develop effective governance that can help ensure that AI-enabled tools are developed and deployed in a responsible and ethical manner. The research objective would involve examining the different approaches to AI governance that have been developed for different industries and identifying the key factors that contribute to the success or failure of these approaches.

The research should focus on a cross-industry deployment, ideally in the healthcare space. The underlying research should examine their current approach to AI governance, including the policies, procedures, and practices that are in place to ensure the responsible and ethical use of AI in their products, including mechanisms in place to ensure that output of the AI can be trusted. Research should review relevant laws, regulations, and ethical guidelines, such as the EU's proposed Artificial Intelligence Act.

Overall, this thesis proposal would provide valuable insights into the challenges of developing effective AI governance frameworks in the selected industry, and the key factors necessary trust the output of AI tools. It would also provide practical recommendations for improving AI governance in a specific company, and potentially contribute to the development of best practices for AI governance.

Background Challenge:

The challenge with AI Governance can be demonstrated with the following example:

“Prompt 1: “what is the main challenge with AI in big organizations”

Answer: One of the main challenges with AI in big organizations is the effective integration and implementation of AI technologies within existing systems and processes. Here are some key aspects that organizations often face:

Data quality and availability: AI systems heavily rely on high-quality and relevant data for accurate decision-making. Big organizations may struggle with disparate data sources, poor data quality, or inadequate data infrastructure, making it difficult to effectively train AI models or extract meaningful insights.

Organizational structure and culture: Implementing AI technologies requires collaboration and coordination across different departments and teams. However, big organizations often have complex hierarchies, siloed structures, and resistance to change, which can hinder the integration and adoption of AI initiatives. Cultural barriers, fear of job displacement, and lack of AI literacy among employees may also pose challenges.

Scalability and resource allocation: Deploying AI at a large scale requires significant computational power, infrastructure, and skilled personnel. Big organizations need to invest in robust hardware, software, and cloud computing resources to handle the computational demands of AI systems. Allocating sufficient resources and securing executive support can be a challenge.

Ethical and legal considerations: AI applications raise ethical concerns regarding privacy, bias, fairness, and accountability. Big organizations must establish clear guidelines and ethical

frameworks for AI development and usage to ensure compliance with legal and regulatory requirements. Implementing proper governance mechanisms to address these concerns can be complex and time-consuming.

Continuous learning and adaptation: AI models need to continuously learn and adapt to changing environments and new data. Big organizations often face challenges in keeping their AI systems up to date, ensuring regular model retraining, and incorporating new information. Staying at the forefront of AI research and advancements is crucial but can be challenging due to resource constraints and other priorities.

User acceptance and transparency: AI solutions may be met with resistance or skepticism from employees and users. It is important to ensure that AI systems are transparent, explainable, and understandable to build trust and user acceptance. Clear communication and education about the capabilities and limitations of AI can help mitigate concerns.

Addressing these challenges requires a strategic approach, involving a combination of technical expertise, organizational change management, data governance, and a commitment to ethical and responsible AI practices."

The above response provided by ChatGPT demonstrates the difficulty in governing/verifying AI tools. Such a detailed response by an AI tool raises questions beyond trust, ethics, rather legal/copyright and oversight. Key questions the research should uncover include:

1. Can organizations *trust* knowledge produced by Large Language Models (ChatGPT) like above?
2. What should governance of using such AI tools include? Mapping out key elements for AI Governance within an organization.
3. How will ethical considerations be integrated into the AI governance framework to ensure responsible and accountable AI use.
4. How can governance be implemented on similar tools across industries, e.g., Healthcare? The aim of the research is to underline steps, procedures and controls that can be implemented to check, verify, and govern the deployment of AI tools as well as the output they produce.

Proposed Research Method:

1. Literature Review: Conduct an extensive literature review to gain a comprehensive understanding of existing AI governance frameworks across various industries (Standard Organizations e.g. ISO, IEEE) . Explore different approaches, best practices, and relevant regulatory guidelines applicable to AI governance in general.

2. Case Study Selection: Select a diverse range of case study companies from different industries or sectors that have implemented AI in their operations. Consider factors such as industry reputation, size, and AI implementation scope to ensure a varied sample.

3. Data Collection: Collect data through interviews, surveys, and document analysis from key stakeholders within the selected case study companies. Explore their current approach to AI governance,

including policies, procedures, and practices in place. Gather information on how ethical considerations are integrated, existing regulatory compliance measures, and any challenges faced.

4. **Data Analysis:** Analyze the collected data using qualitative research methods such as thematic analysis. Identify common themes, patterns, and key insights related to the success or failure of AI governance approaches across different industries. Compare and contrast findings to extract overarching principles and factors contributing to effective AI governance.

5. **Recommendations:** Based on the analysis, develop a gap-analysis with existing AI regulations and create a set of harmonized recommendations for improving AI governance frameworks applicable to cross-industries. Propose high-level changes to policies, procedures, and practices that can enhance responsible and ethical AI use in different organizational contexts. Focus on providing actionable strategies that can be customized based on industry-specific requirements.

6. **Validation and Feedback:** Validate the proposed recommendations by seeking feedback from experts in the field of AI governance and industry practitioners. Incorporate their input and suggestions to enhance the relevance and applicability of the recommendations.

7. **Report and Presentation:** Compile the research findings, recommendations, and insights into a comprehensive report. Prepare a presentation summarizing the research process, key findings, and proposed generic recommendations. Present the findings to relevant stakeholders, including industry professionals, policymakers, and academia, to promote knowledge sharing and further discussion.

Required Outcomes:

Based on this analysis, the outcome of the research should propose a set of recommendations for improving AI governance framework, including any necessary changes to policies, procedures, or practices in the given industry. The student would also identify any gaps or weaknesses in the existing approach and propose strategies for addressing these issues. Consequently, the following AI Governance Framework practices should be mapped:

- **Context:** A mapping of the business environment where AI is deployed, as well as key stakeholders involved. Context should also consider if the AI tool is developed in-house or provided by a 3rd party vendor.
- **Regulation:** An overview of regulations relevant to the industry/company regarding use of AI (required reading: ISO/IEC 42001 Standard).
- **Planning/Deployment:** Steps involved in integration of AI tools within an organization
- **Operationalization:** After deployment, how are KPIs set and determined? Who is responsible for oversight? (Recommended reading AI HLEG guidance document).
- **Ethical:** What measures must be set, implemented to ensure ethical usage of the AI? (How is an ethical review board determined).

Required Reading:

https://mitsloan.mit.edu/ideas-made-to-matter/making-most-ai-latest-lessons-mit-sloan-management-review?utm_source=mitsloanlinkedin&utm_medium=social&utm_campaign=smrq22023

This research is part of a larger doctoral research on AI Assessment with TÜV SÜD.

Contact: Interested master thesis students should send documents (CV and relevant work) to:

Edmund Balogun

edmund.balogun@tuvsud.com

Dr. Ibrahim Halfaoui

ibrahim.halfaoui@tuvsud.com