ITEA Topical roadshow
Large Language Models

Engineering considerations in LLM governance and trustworthiness

26 March 2024 | Online

Mikko Raatikainen (University of Helsinki)

Engineering considerations in LLM governance and trustworthiness

Mikko Raatikainen
University of Helsinki

https://researchportal.helsinki.fi/en/persons/mikko-raatikainen

My background

University Researcher, Title of Docent, D.Sc (eng.)

- Software engineering: Software architecture, requirements engineering, and software business.
- Long and broad experience in research-industry collaboration.

UNIVERSITY OF HELSINKI

IML4E

ITEA projects:

- Industrial Grade Machine Learning for Enterprises (IML4E). (5/2021-9/2024). www.iml4e.org
- Engineering large foundational models (ELFMo), ITEA labeled (10/2024-9/2027)

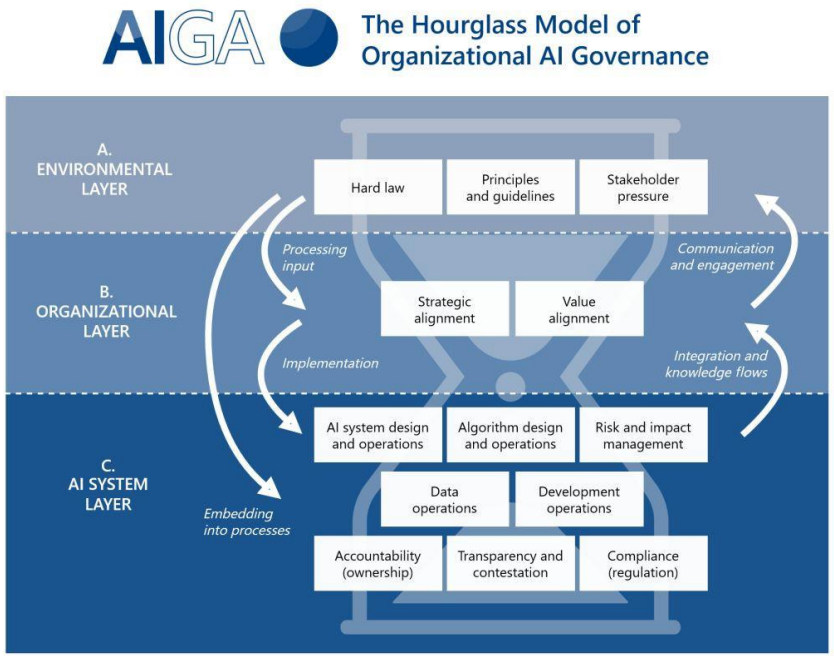
AI governance, AI ethics, AI regulation…

High societal and authority interests.

- GDPR, AI Act.
- Sustainability (/ESG), etc.
- Domain/organizational specifics.

→ Conceptual knowledge and high-level practices and frameworks have emerged.

→ Minor impact in engineering or practical tasks (surveys in Finland).



AI Act vs GDPR vs … ?

Which one will have the biggest impact?

What will be the practical implications for engineers?

How do LLMs change the landscape?

Earlier everyone were talking about AI...

...but now everyone are talking about how they use chatGPT.

→ LLMs have made understanding and application of AI easier.

Pros: High opportunity for novel services, even disruptive designs.

Cons: Technology unawareness elevates risks, trustworthiness concerns, and ethical issues.

AI lineage

A ***holistic and integrated information framework*** designed to handle **integrated information** throughout the lifecycle, especially differentiating

- ***Development (model-level)*** *and*
- ***Operational (prediction level)***

Facilitates the comprehensive and holistic capture of ***various concerns***, including those related to business, compliance, quality, and development information.

Provide stakeholders with a tool to navigate the entire ***life cycle journey.***

Take into account and integrate the ***existing tools*** and their information.

AI lineage in today's workflows and pipelines

Engineering workflows and DevOps/MLOps pipelines can relatively effortlessly generate and capture a significant portion of the information.

- E.g., Issue tracking for downstream workflows with quality gates, MLFlow for ML experiments, prometheus for monitoring, model cards, etc.

However, **challenges for LLMs**

- Workflows not fully established
- Information is not integrated across tools.
- High-level measures, such as KPIs, QA, or ethics.
- Prediction level information is limited, such as for prompt engineering.

Nevertheless, a minimal AI lineage could be constructed with relative ease but requires pro-active work.

Summary

LLMs facilitates service design, but also requires engineering rigor

→ The great power of LLMs and GenAI comes with great responsibility

Governance and trustworthiness requires AI lineage solution covering LLM-based service development and operations.