Governing AI with Trias Politica

B. van Gils and J. B. F. Mulder

Antwerp Management School (AMS)

Author Note

The idea for this thesis project originated in a discussion with Prof. B. Derksen and the two authors of this proposal. We decided to write the proposal with just two authors. Of course, part of the 'kudos' are for Prof. Derksen.

Abstract

Artificial Intelligence (AI) is literally everywhere. It helps us in our daily lives in the 'smart' appliances that we use, it helps us to find the most optimal route through traffic, and it helps in numerous business tasks. AI remains a means to an end. Over-use of AI may lead to disastrous results, both in the public domain and for organizations. We believe that AI-governance is needed. This paper describes a research proposal to use the ideas behind the Trias Politica and allow AI-systems to govern AI-systems in a business context.

Keywords: Digital transformation, AI, Governance, Architecture

Governing AI with Trias Politica

Introduction

Artificial Intelligence (AI) tools are widely used in various contexts. To illustrate this, consider the following. In our personal lives, we likely encounter AI on our smartphone (e.g., SIRI), in the software that powers our doorbell/home security system, in the smart appliances in our home, the navigation system in our car, or even as the driver of our car. In a business context, the use of AI goes even further than that. We see the use of AI as tool for quickly researching innovative ideas, for helping to draft E-mails, to create visuals for presentations, to write code, and even to make business decisions (Chuma & de Oliveira, 2023).

The use of AI certainly has great potential (benefits) but also comes with serious risks (costs) (Parnas, 2017; Steimers & Schneider, 2022). An example in our personal context is the use of the self-driving function of a car. Suppose the AI mishandles a situation and causes you to crash, causing huge damage to other vehicles. Do you really *a*) have that on your conscience, and *b*) be liable for the associated costs? The same applies to a business context. Suppose the AI mishandles a situation and treats a customer unfairly (e.g. racial profiling) or writes bad code. The question of ethics and liability once again apply.

Based on this (over simplified) expose, we conclude that the unchecked use of AI is probably a bad idea. This claim is supported by the fact that the European Union (EU) has recently published it *AI-act* (European Parliament and Council of the European Union, 2023). The AI Act is a legal framework governing the development and use of artificial intelligence in the EU. We expect this *act* to evolve in years to come to accommodate for the new status quo as it emerges. With legislation in place, we can at least (objectively) decide what is/should be allowed when using AI in the EU.

The same line of reasoning applies also on the smaller scale of a company (or

perhaps even at the level of a business ecosystem). The use of AI in companies should also be *governed*. This leads to the notion of *AI Governance* (Birkstedt et al., 2023) which appears to be closely related to other governance disciplines such as *IT Governance* and *Data Governance* (De Haes et al., 2020; Hoogervorst, 2017; Van Gils, 2023).

There are several important questions to consider in this context. First and foremost: Who or what does the governing? Would it be humans that govern the AI? Or can one AI govern another AI? Or perhaps it should be a collaboration of a human and AI doing the governing together. A second important question is: What type of governance mechanisms would we be able to use in such a context? For example, from the work of De Haes et al. (2020), we know that there are structure, relational, and process mechanisms. Which of these apply here? A third question is more foundational and philosophical in nature and pertains to the 'trias politica' (also known as 'the separation of powers'[1]). Simply put, this is about separating *a*) the creation/definition of laws, from *b*) executing them, and *c*) the judicial system for handling (potential) violations of the law. The question that arises is: would it be a good idea to apply these principles also for AI-governance, and what would that look like?

Research topic

We propose to research the questions around AI-governance as mentioned in the previous section. We are aware that this is a big topic and therefore we suggest to tackle these questions with a dual-thesis or even triple-thesis. Collaboration with peers will help to make good speed and cover much more ground. Our experience is also that it is more effective (students can help each other) and more fun.

Various angles and approaches are possible for this research. As a starting point, we give the following suggestions (but strongly invite you to tune these to your own

[1] https://en.wikipedia.org/wiki/Separation_of_powers (last checked: 02-jan-2024).

liking):

- Taking an ontological and philosophical perspective, you can zoom in on the question: what *is* AI-governance exactly. Such an approach would require you to do a significant literature study for sufficient academic rigor. To tackle practical relevance, you could work with experts to verify your definition. Even more, you could do a case study to see if AI-governance in a real organization meets the criteria of your definition.

- Taking a design science perspective, you build an AI-governance framework while taking the principles of the Trias Politica into account. Requirements for such an artefact may come from literature as well as experts. Once you define your artefact, you can test it using an expert session.

- Taking an experimental approach–likely using design science as an underlying methodology–you can not only define, but also build a working system that implements a governance framework. As 'food for thought', imagine using OpenAI's ChatGPT as the legislator, Microsoft's Copilot as the executing power, and Google's Bard as a judicial power.

Depending on the angle that you take in your work, we can recommend several sources for helping you to get started. A preliminary list:

- If you want to use group support systems to quickly collect data from groups of experts, then your promotor can help you with Meetingwizard. If you need more in-depth support, ask help from Prof. Mulder, Prof. Bobbert, or Prof. Van Gils.

- We have several experts on governance on staff that you can ask for help, e.g. Prof. De Haes and Prof. Derksen.

- If you want to do an experiment, or talk to people who use AI in their data to day work, then we can suggest various companies for you to work with. For example, we have good connections with Mavim that we can use to this end.
- If you are interested in the Security or the Data/Data Management aspects–which certainly are relevant for this research–then we recommend you to work with e.g. Prof. Bobbert or Prof. Van Gils respectively.

Parting thoughts

The slogan and mindset for research at AMS is "opening minds to impact the world". We believe that the topic of AI and AI-governance is highly relevant in today's society. It bridge the domains of business and IT and gives answers to questions that we (will) face in society (in the near future). It is a topic that hasn't seen much academic research yet. Therefore, we suspect that your thesis will not only be exciting, but also ground-breaking.

We hope that you are enthusiastic about the topic and encourage you to think about it. You are more than invited to get in touch to discuss this further.

References

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. *Internet Research*, *33*(7), 133–167.

Chuma, E. L., & de Oliveira, G. G. (2023). Generative AI for business decision-making: A case of ChatGPT. *Management Science and Business Decisions*, *3*(1), 5–11.

De Haes, S., Van Grembergen, W., Joshi, A., & Huygh, T. (2020). *Enterprise governance of information technology* (Vol. 3). Springer.

European Parliament and Council of the European Union. (2023). Artificial intelligence act [https://artificialintelligenceact.com/ (last checked 03-jan-2024)].

Hoogervorst, J. A. P. (2017). *Foundations of enterprise governance and enterprise engineering. Presenting the employee-centric theory of organization*. Springer.

Parnas, D. L. (2017). The real risks of artificial intelligence. *Communications of the ACM*, *60*(10), 27–31.

Steimers, A., & Schneider, M. (2022). Sources of risk of AI systems. *International Journal of Environmental Research and Public Health*, *19*(6), 3641.

Van Gils, B. (2023). *Data in context. Models as enablers for managing and using data*. Springer.