



Can AI go to Jail? – Consequences for the Corporate Governance

“Success in creating effective AI, could be the biggest event in the history of our civilisation. Or the worst. We just don't know”
Stephen Hawking as cited by Girasa (2020, p.64)

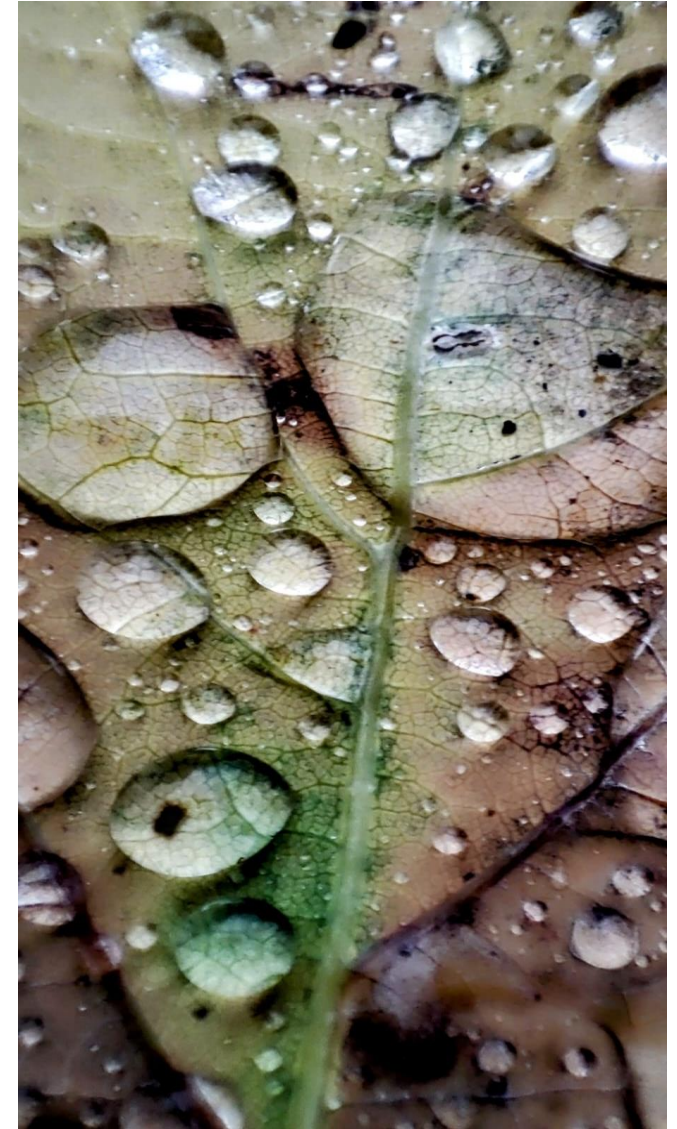
EAA e-Conference on
Data Science & Data Ethics

16 May 2023

Steef Peters
Gerrit Jan van den Brink
The Hague University for Applied Sciences

AGENDA

- View on learning AI and the related issues
- How to cope with the issues
- Areas for further research



VIEW ON LEARNING AI AND THE RELATED ISSUES



VIEWS

- Legal:
 - Humans remain accountable for all the Artificial Intelligence they create
- Regulatory/Supervisory
 - All created AI needs to be explainable and non-discriminatory (Elderson, ECB; DNB)
- Ethical
 - Humans have to stay in control (see Huber, 2022, p 125, p 134)
 - Humans should have control over the „stop“ button: Autonomy

Are these starting points?

TAEIHAGH DESCRIBES THE RISK OF PROGRESSING AI (P 140 -143, EXTRACT)

- AI can face so called “corner cases” (areas with learning problems for the AI system) causing unintended and unwanted results
- Autonomy in decision making by AI can significantly reduce human autonomy
- Learning AI systems can potentially exhibit behaviours that conflict with societal values and norms

HOW TO COPE WITH THE ISSUES



USING AI GOVERNANCE

AI governance is a system of rules, practices, processes and technological tools that are employed to ensure the use of AI technologies according to the objectives and values, according to the legal requirements and meeting the principles of ethical AI

(M. Mäntimäki e.a., 2022, p 2)

AI GOVERNANCE SHOULD BE INTERNALISED

GIVEN THE DEVELOPMENT PACE AI SHOULD CONTROL ITSELF TO THE LARGEST AMOUNT POSSIBLE

- Business and communities should be prepared for a quicker independent development of AI
- As a consequence human control can easily deteriorate and might only be exercised indirectly in the (near) future
- Therefore we need to build AI embedded risk management and monitoring tools, spotting areas in which human ethical principles are being neglected or even violated

GOVERNANCE IN MULTI OBJECT NETWORKS BASED ON FIELD THEORY

- In this research the authors explore potential solutions for AI governance. The question to be answered here is if AI can be run with an "included conscience" based on defined ethical AI standards, which is able to develop itself quick enough to protect organizations against detrimental effects?
- We will explore a framework which is based on the Network Field Model. This model was borrowed from the physics science.
- In the research we link AI systems next to the humans as objects in a network and use the trust as value of the social capital involved to determine the governance in the network.
- The authors are aware of the size of the problem. Therefore the authors plan a second part which will use the developed framework in further research in this area.

NETWORK FIELD MODEL VS SYSTEM MODEL

- System model
 - Interaction between two objects
 - Time is a parameter
 - Structure is hierarchy

- Field model
 - Influence by the total field on all objects
 - Time is a coordinate
 - Structure is heterarchy (layers of networks and hierarchy)

FUTURE SCOPE

Features	Hierachy	Network
Organizing metaphor	Highways	Surface streets
Organizing principles	Depth, breadth, inverse relationship	Centrality: hub, gatekeeper and pulsetaker
Protocol	Formal	Informal
Elasticity	Rigid	Flexible
Visibility	Explicit	Implicit
Relationship	Authority	Trust
Diversity	Heterogeneous	Homogeneous
Knowledge stored in	Procedures	Relationships
Half-life	Centuries	Generations
Power	Direct: command and control	Indirect: influence and mutuality
Resource investment	Financial capital	Human and social capital

AREAS FOR FURTHER RESEARCH



GOVERNANCE GOVERNING THE CONSCIENCE

- Integrate legal, ethical and regulatory aspects in AI systems
- AI systems should protect themselves for decisions and further development outside of the ethical principles and legal requirements
- Until the full own protection human control needs to be in place
- Sanctions are penalties based on the universal rights of humans for autonomy

SANCTIONS NEED TO BE IN PRINCIPLE EXERCISED BY AI EMBEDDED MODULES

- Various sanctions could help keeping AI developments straight:
 - Isolate a malfunctioning AI component
 - Adjust the AI principles (with human intervention)
 - Put AI component under administration
 - Delete the AI component
- AI is in principle operating and the network-level, not at the hierarchy level
- The question regarding AI operating on the hierarchy level for the isolation component needs to be further researched
- Sanctions need to be taken very carefully since other AI components might rely on the component which is to be sanctioned. All sanctions except for adjusting the AI principles could have severe unintended consequences

Steef Peters

- PhD in elementary particle physics
- CIO leasing company
- Startups in DeFi, Global Outsourcing
- Professor in Strategy and Business Informatics

Gerrit Jan van den Brink

- PhD in economics, Chartered Accountant
- CRO insurance company
- Startups in modern consulting businesses
- Lecturer on Risk Management, Cyber Security, ESG, Inflation

ABOUT US



Entrepreneurial
experimentalist /
wallet79



Founder and MD Risk
Sigma GmbH

REFERENCES

- Cihon, P., Schuett, J., & Baum, S. B. (2021, July 5). *Corporate Governance of Artificial Intelligence in the Public Interest*. Retrieved September 2022, from Information 2021: <https://doi.org/10.3390/info12070275>
- Evrin, T., Stanislav, M., & Joep, C. (2022). *Blockchain governance in the public sector: A conceptual framework for public management*. Retrieved 8 2022, from www.sciencedirect.com: <https://doi.org/10.1016/j.giq.2021.101625>
- Hilb, M. (2020, July 21). Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance. *Journal of Management and Governance*, pp. 851-870.
- Huber, W. (2022). *Menschen, Götter und Maschinen, Eine Ethik der Digitalisierung*. München: C.H. Beck.
- Mäntymäki, M., Minkinen, M., Birkstedt, T., & Viljanen, M. (2022, Feb. 24). *Defining organisational AI governance*. Retrieved September 2022, from AI and Ethics: <https://doi.org/10.1007/s43681-022-00143-x>
- Peters, S., & Stephenson, K. (2022). *Toward a General Theory of Organizing, Volume 1: Introducing the Network Field Model*. Retrieved from <http://dx.doi.org/10.5772/intechopen.99709>
- Renda, A. (2019). *Artificial Intelligence. Ethics, governance and policy challenges*. Brussels: Center for European Policy Studies.
- Taeihagh, A. (2021). *Governance of artificial intelligence*. Retrieved September 2022, from <https://doi.org/10.1080/14494035.2021.1928377>



Thank you very much for your attention

Contact

Gerrit Jan van den Brink

Risk Sigma GmbH

gjvdbrink@risksigma.de

+49.176.53788522

EAA e-Conference on
Data Science & Data Ethics

16 May 2023