

**FUTURES**  
W/OUT WORK

# RESPONSIBLE AI

Aleksandre Asatiani  
*Department of Applied IT*  
*University of Gothenburg*



UNIVERSITY OF  
GOTHENBURG

# Aleksandre (Alex) Asatiani



- Associate Professor in *Information Systems*, University of Gothenburg
  - Previously: Aston University (UK), Aalto University (Finland)
- **Teaching topics:**
  - Digital Strategy
  - Digital Platforms
  - Future of Work
- **Research interests:**
  - Digital transformation of work
  - Emerging tech & organizational processes

# Organizational/Corporate Responsibility



# Organizational/Corporate Responsibility



**FUTURES  
W/OUT WORK**

# Responsibility domains

- Legal accountability: regulations explicitly prohibit irresponsible actions:
  - Child labor
  - Unsafe work environment
  - Use of certain chemicals in food
- Ethics: not illegal, but questionable
  - Marketing to vulnerable populations
  - Addictive user interfaces (contested currently)
  - Less clear-cut bias in decisions

# Responsibility and (Information) Technology

- Technology (often provided by a third party) have big role in organizational activities
- Distributed responsibility
- Impact of design decisions



# Technology is not neutral

- Counter argument to:
  - Technological determinism
  - Neutral technology
- Deliberate design choices influence technology use
- These design choices are *political*
- Hence - consequences of technology use are political



LANGDON WINNER

Do Artifacts Have Politics?

# Loan Decision



**SHOULD THE BANK  
IMPLEMENT AI FOR LOAN  
DECISIONS?**

# Key question for this lecture:

*How can organizations benefit from AI without sacrificing accountability, fairness, human judgment, and trust?*

# Responsible AI



*Responsible AI is the practice of developing, using and governing AI in a human-centred way to ensure that AI is worthy of being trusted and adheres to fundamental human values.*

**Source:** Vassilakopoulou, P., Parmiggiani, E., Shollo, A., & Grisot, M. (2022). Responsible AI: Concepts, critical perspectives and an Information Systems research agenda. *Scandinavian Journal of Information Systems*, 34(2), 3.

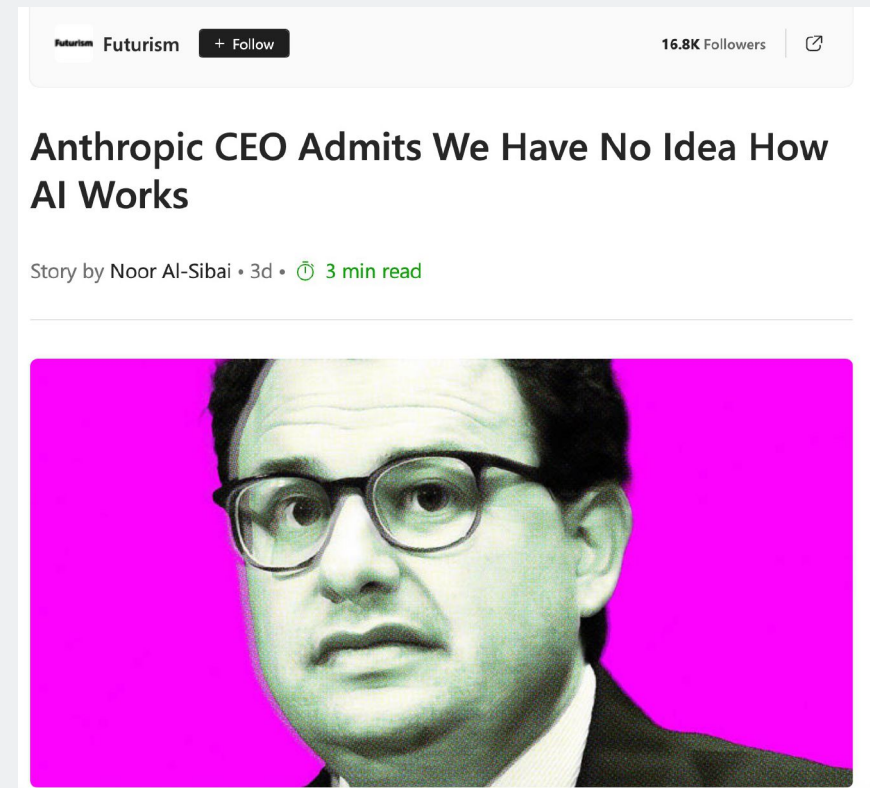
# Layers of Responsible AI



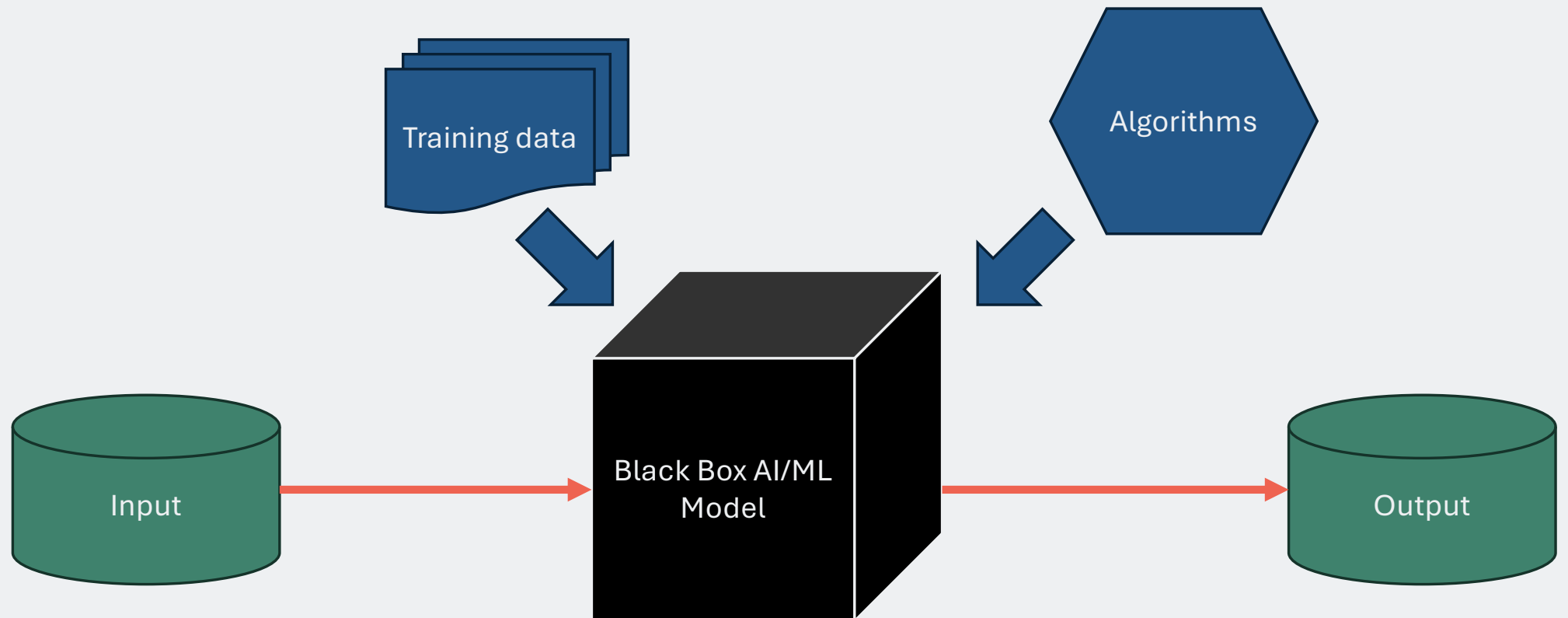
# **Some Pressing Challenges**

# Black-box AI models

- AI models that are hard or impossible to scrutinize
- Lack of transparency, interpretability, explainability

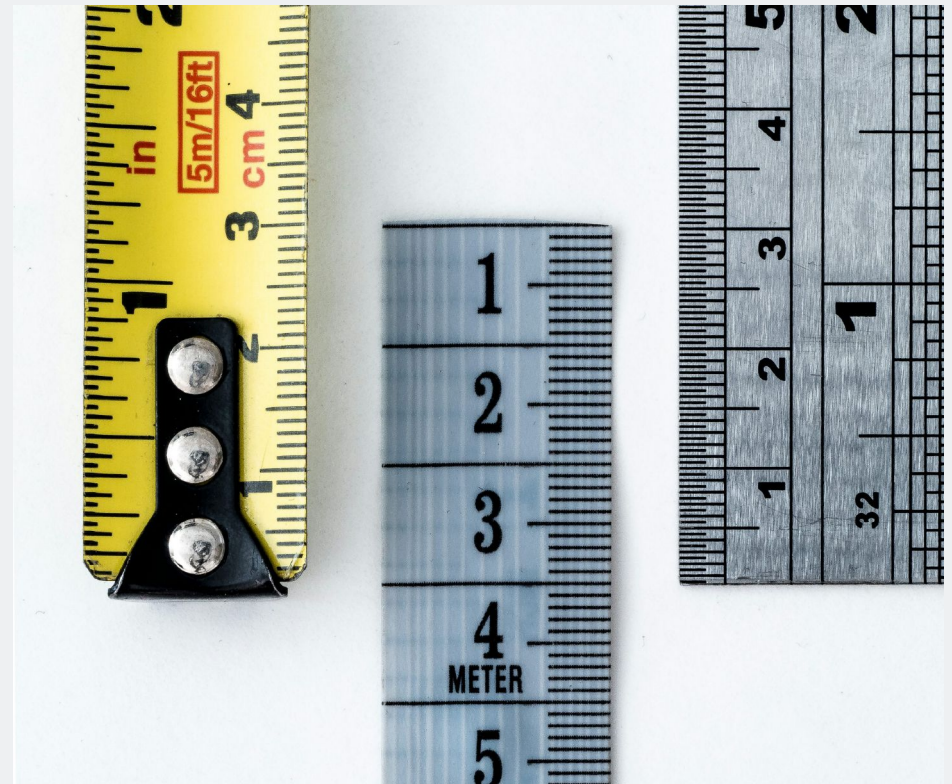


# Black-box AI models



# Training Data Problems

- Bias and discrimination in training data
  - Garbage in – Garbage out
- Invisible and evolving biases
- Context transferability



# Privacy and Surveillance

- Greater accuracy requires more data
- What is an acceptable trade-off between accuracy and privacy
- Inability (lack of incentives) to ensure data safety



# Legal Accountability Gaps

- Legislative lag in fast-developing environment
- Lobbying
- Impossibility of legislating everything



# Cost-Benefit for AI Operators

- Gains of irresponsible behavior can surpass potential downsides
- Increasing downsides (e.g., legislatively) can damage development



**FUTURES**  
W/OUT WORK

# Ethical AI Principles to the Rescue

|   |                                                                         |
|---|-------------------------------------------------------------------------|
| 1 | Benevolence: Do good                                                    |
| 2 | Non-maleficence: Do no harm                                             |
| 3 | Autonomy: Safeguard meaningful human control                            |
| 4 | Justice: Promote fairness, equality, unbiasedness                       |
| 5 | Explicability: Ensure transparency, interpretability and accountability |

# Ethical Principles are Hard in Practice

- Translating a principle into *concrete* action
- Productively working around AI system complexity and limitations
- Messiness of organization
- Distributing responsibility across people, processes, vendors, users, and institutions.



# Examples of Thing Going Wrong

**Robo-debt class action could deliver justice for tens of thousands of Australians instead of mere hundreds**

September 17, 2019 12:28pm BST

Labor's government services spokesman Bill Shorten and lawyer Peter Gordon an

Email

Twitter

14

Facebook

5.9k

LinkedIn

The announcement by Gordon compensate victims of the gove scheme is welcome, perhaps ev

LAW.gov

LIBRARY  
OF CONGRESS

LAW

Search this site

GO

Law Library Home

Global Legal Monitor



[Home](#) | [Search](#) | [Browse All Jurisdictions](#) | [Browse All Topics](#) | [Browse All Authors](#) | [RSS](#) | [Top Recent Articles](#)

## Netherlands: Court Prohibits Government's Use of AI Software to Detect Welfare Fraud

(Mar. 13, 2020) On February 5, 2020, the District Court of The Hague (Rechtbank Den Haag) [held](#) that the System Risk Indication (SyRI) algorithm system, a legal instrument that the Dutch government uses to detect fraud in areas such as benefits, allowances, and taxes, violates article 8 of the [European Convention on Human Rights \(ECHR\)](#) (right to respect for private and family life).

# Any Solution Needs to be Socio-technical

Technical design + Organizational practices

# AI Envelopment



ISSN 1536-9323

Journal of the Association for Information Systems (2021) 22(2), 325-352  
doi: 10.17705/1jais.00664

RESEARCH ARTICLE

Copyright owned by Association for Information Systems. Use for profit is not allowed.  
Link to the Journal of AIS: <https://aisel.aisnet.org/jais/>

## Sociotechnical Envelopment of Artificial Intelligence: An Approach to Organizational Deployment of Inscrutable Artificial Intelligence Systems

**Aleksandre Asatiani<sup>1</sup>, Pekka Malo<sup>2</sup>, Per Rådberg Nagbøl<sup>3</sup>,  
Esko Penttinen<sup>4</sup>, Tapani Rinta-Kahila<sup>5</sup>, Antti Salovaara<sup>6</sup>**

<sup>1</sup>University of Gothenburg, Sweden, [aleksandre.asatiani@ait.gu.se](mailto:aleksandre.asatiani@ait.gu.se)

<sup>2</sup>Aalto University School of Business, Finland, [pekka.malo@aalto.fi](mailto:pekka.malo@aalto.fi)

<sup>3</sup>IT University of Copenhagen, Denmark, [pena@itu.dk](mailto:pena@itu.dk)

<sup>4</sup>Aalto University School of Business, Finland, [esko.penttinen@aalto.fi](mailto:esko.penttinen@aalto.fi)

<sup>5</sup>The University of Queensland, Australia, [t.rintakahila@uq.edu.au](mailto:t.rintakahila@uq.edu.au)

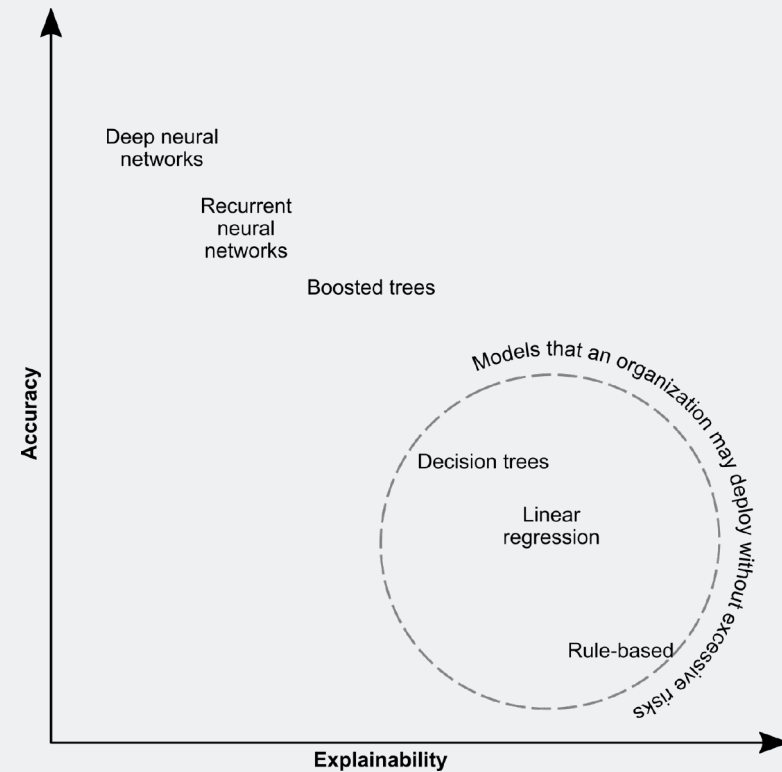
<sup>6</sup>Aalto University School of Arts, Design and Architecture, Finland, [antti.salovaara@aalto.fi](mailto:antti.salovaara@aalto.fi)

### Abstract

The paper presents an approach for implementing inscrutable (i.e., nonexplainable) artificial intelligence (AI) such as neural networks in an accountable and safe manner in organizational settings. Drawing on an exploratory case study and the recently proposed concept of envelopment, it describes a case of an organization successfully “enveloping” its AI solutions to balance the performance benefits of flexible AI models with the risks that inscrutable models can entail. The authors present several envelopment methods—establishing clear boundaries within which the AI is to interact with its surroundings, choosing and curating the training data well, and appropriately managing input and output sources—alongside their influence on the choice of AI models within the organization. This work makes two key contributions: It introduces the concept of sociotechnical envelopment by demonstrating the ways in which an organization’s successful AI envelopment depends on the interaction of social and technical factors, thus extending the literature’s focus beyond purely technical issues. Secondly, the empirical example illustrates how organizational factors

# Why use the black box then?

Black-box AI is mysterious,  
but powerful



# Potential solutions

Make AI explainable VS. Envelop AI

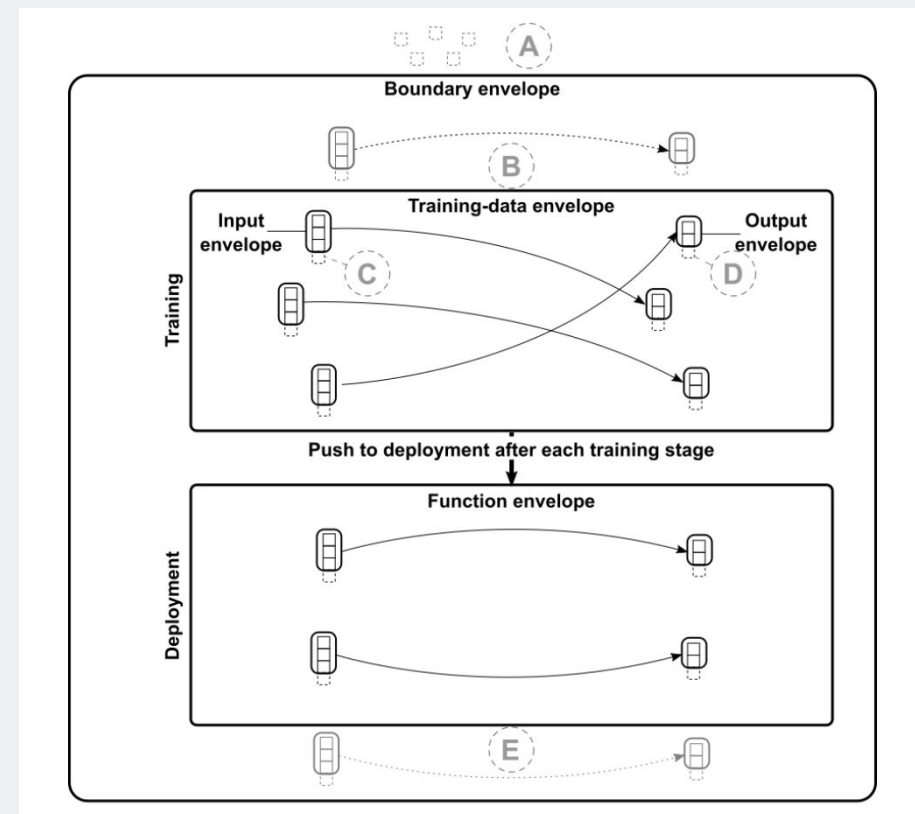
# Envelopment in robotics



# Envelopment by Scott Robbins

1. Boundary envelope
2. Training-data envelope
3. Input envelope
4. Output envelope
5. Function envelope

Robbins' discussion is largely 'tech' focused



Robbins, S. (2020). AI and the path to envelopment: knowledge as a first step towards the responsible regulation and use of AI-powered machines. *AI & Society*, 35(2), 391-400.

Floridi, L. (2011). Children of the fourth revolution. *Philosophy & Technology*, 24, 227-232.

# Case study

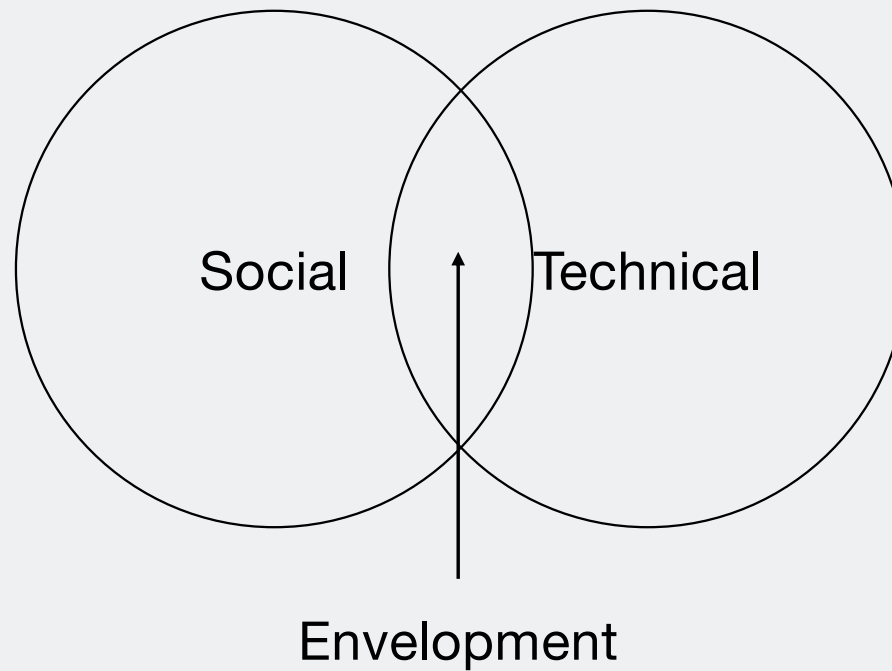
- Danish Business Authority (DBA) – an authority charged with a wide variety of tasks related to companies and business development in Denmark
- Tasked with maintenance and enforcement of Danish Companies Act, Financial Statements Act, Bookkeeping Act, and Act on Commercial Foundations

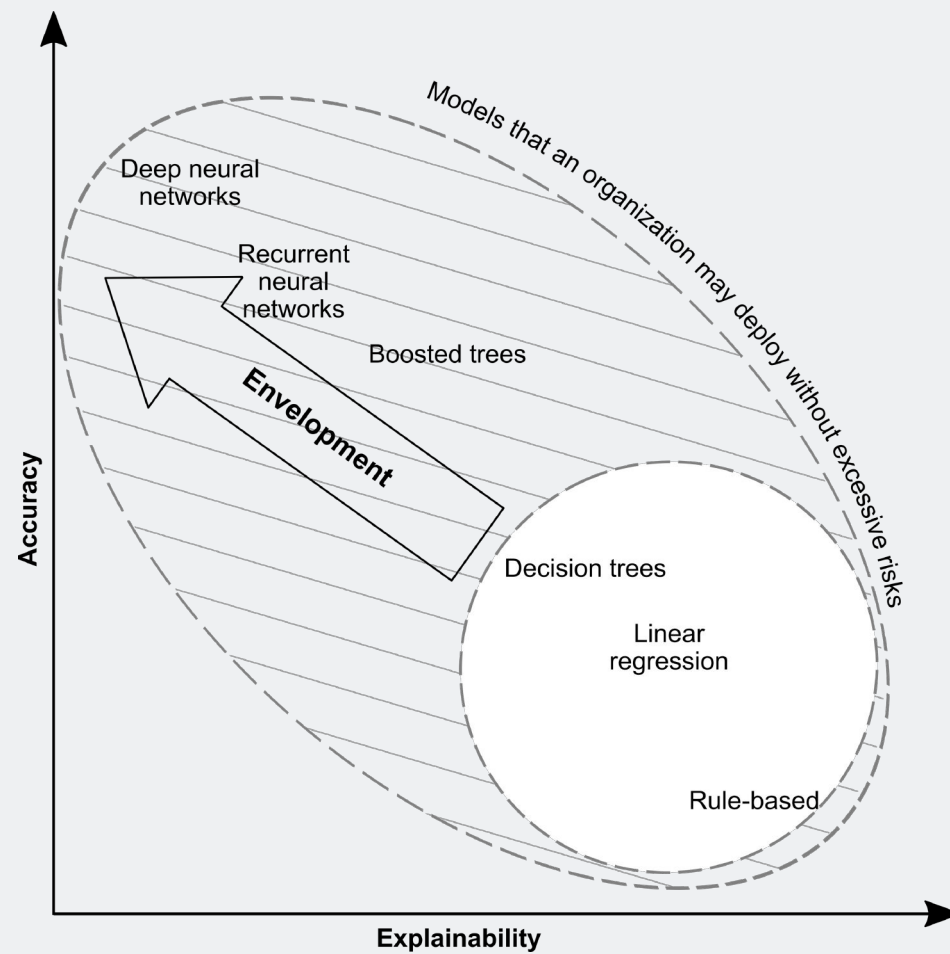


# DBA's goals

- Improve efficiency of the authority, service quality, and work environment for case workers.
- AI/ML a priority.
- Very important not to tarnish image and reputation of Danish public authorities.

**Envelopment has to happen at the intersection of social and technical**





# Implications

- Both instrumental and humanistic outcomes need to be prioritized.
- Parties impacted by and using the AI system need to participate in design and envelopment.
- Full control should not be given to AI.
- Simple is the best.

**THANK YOU!**

**QUESTIONS?**



UNIVERSITY OF  
GOTHENBURG

**FUTURES  
W/OUT WORK**