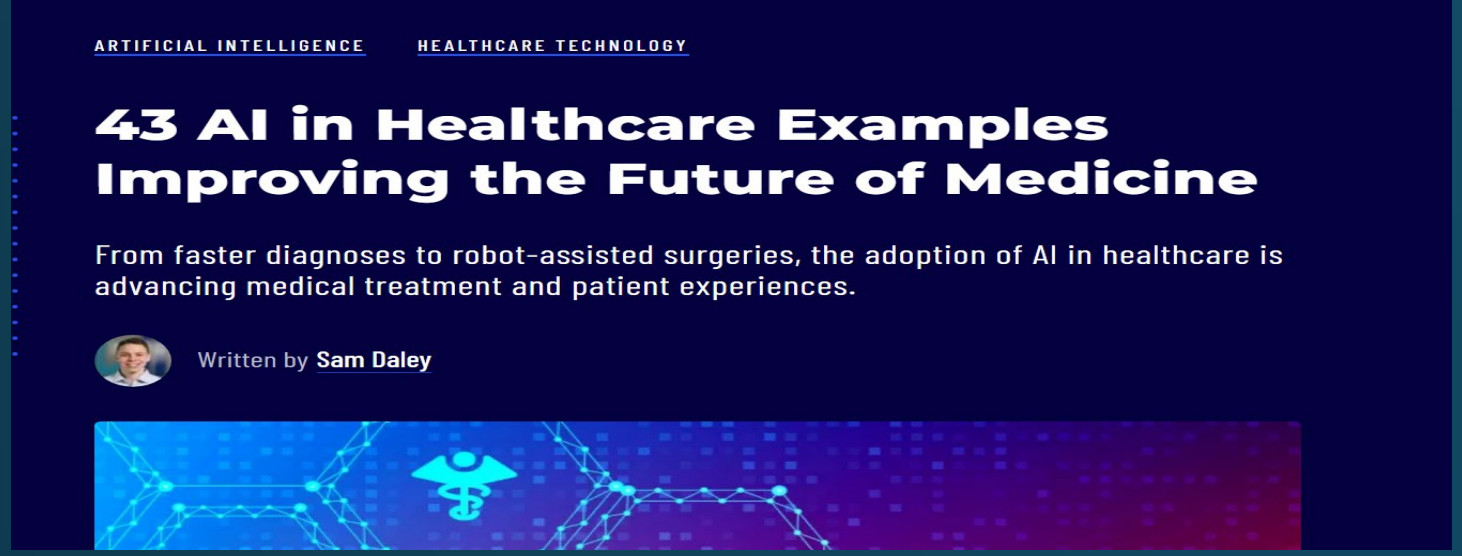
AI for Healthcare

Tech Collaboration Lab discussion

Jeroen van de Hoven and Stefan Buijsman

13-6-2025



Innovate Responsibly with AI for Health & digital ethics for health: Medical Ethics 2.0

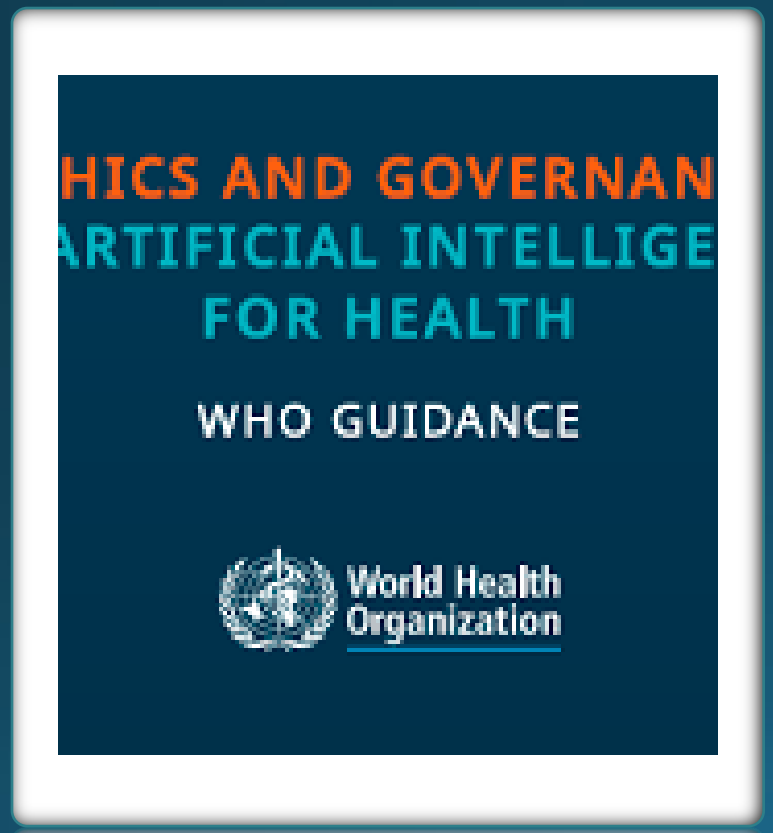
WHO: Ethics and AI (2021)

WHO has adopted it as preferred approach to digital Ethics

- >>. AI systems should be designed demonstrably and systematically to conform to the principles and human rights with which they cohere; more specifically, they should be designed to assist humans, whether they be medical providers or patients, in making informed decisions.<< (p. 25)

TU Delft

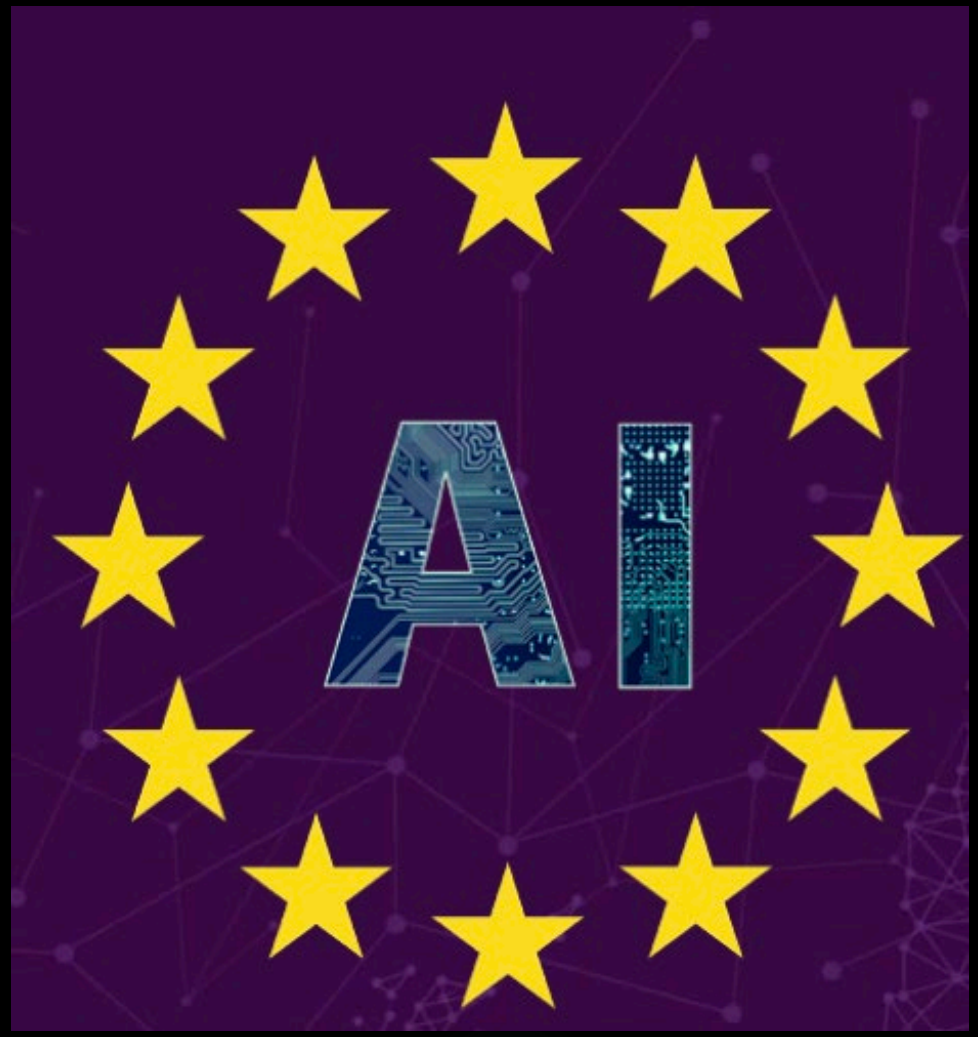
Design for Values



- >>"Design for values" is explicit transposition of moral and social values into context-dependent design requirements. (…) a roadmap for stakeholders to translate human rights into context-dependent design requirements through a structured, inclusive, transparent process, such that abstract values are translated into design requirements and norms (properties that a technology should have to ensure certain values), and the norms then become a socio-technical design requirement. << (p. 66)

Aligned with EU Values for AI

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability



European Foundations

An approach to AI Ethics that is gaining ground: "Think Design"

Specification, Operationalization in terms of requirements

Does not replace critical moral analysis, ethical debate, philosophical reflection

Paul Nemitz on AI

- In order to protect and strengthen Western Liberal democracies in the Age of AI and the core trinitarian idea of 'human rights, rule of law and democracy' we need " a new culture of technology and business development …which we call human rights, rule of law and democracy by design"


PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

You have access

Check for updates

MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

View PDF

Tools Share

Cite this article

Section

Abstract

1. Introduction

2. A unique concentration of power in the hands of digital Internet giants which develop artificial

Opinion piece

Constitutional democracy and technology in the age of artificial intelligence

Paul Nemitz

Published: 15 October 2018 https://doi.org/10.1098/rsta.2018.0089

Abstract

Given the foreseeable pervasiveness of artificial intelligence (AI) in modern societies, it is legitimate and necessary to ask the question how this new technology must be shaped to support the maintenance and strengthening of constitutional democracy. This paper

EU VALUES BY DESIGN WITH AI

VALUE/ DESIGN NEXUS

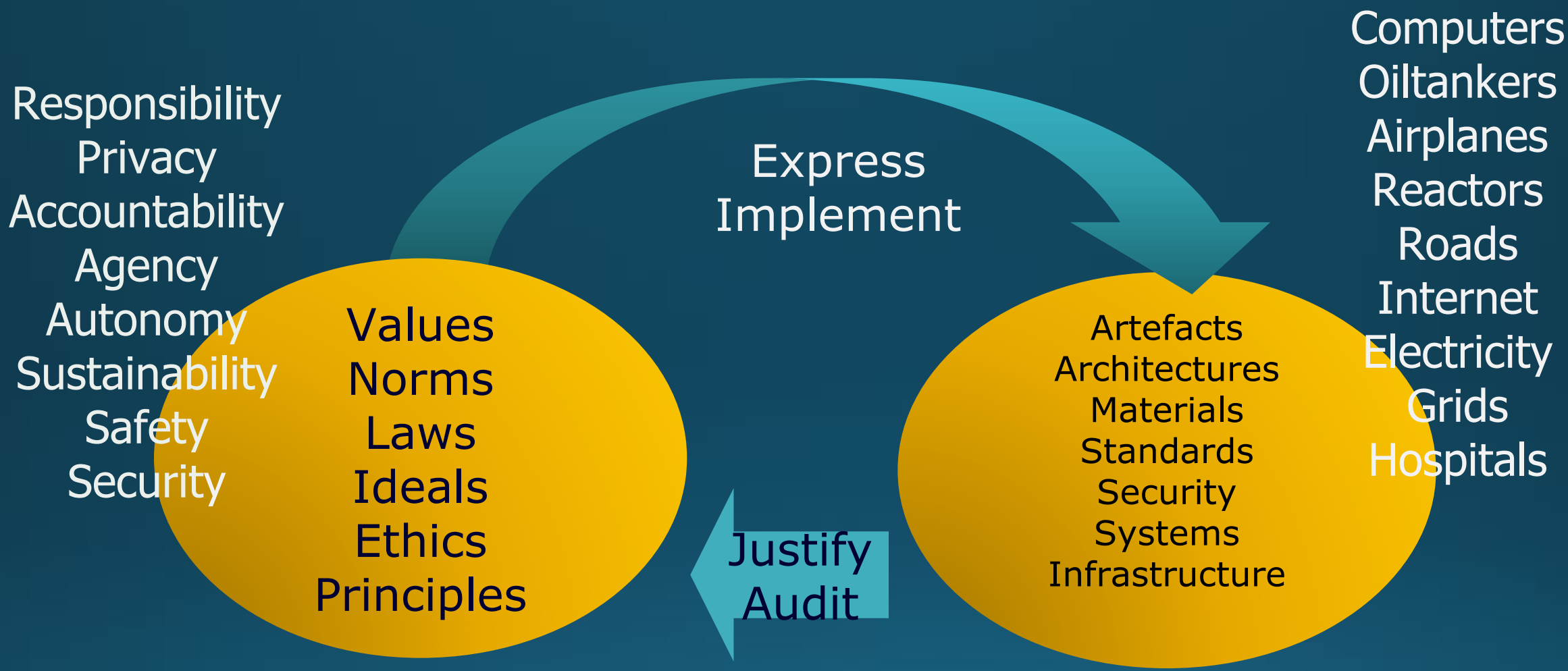
IF WE DO NOT DESIGN FOR WHAT MATTERS: IT WILL NOT HAVE A PLACE IN TOMORROWS WORLD

Designs are value laden

Values are Design Consequential

Racism is baked in healthcare AI

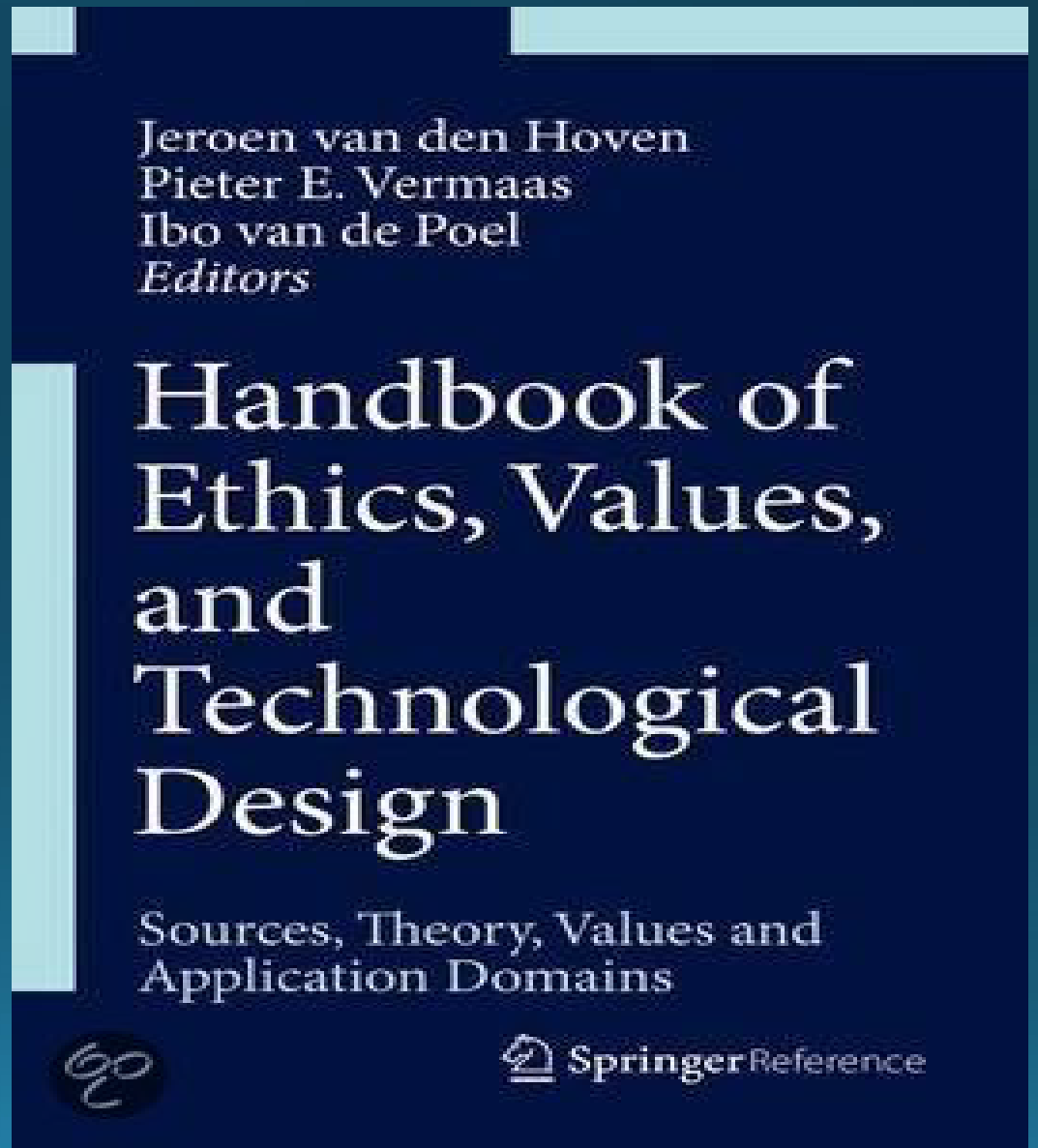
Algorithm gave a sicker black person the same health risk score as a healthier white person



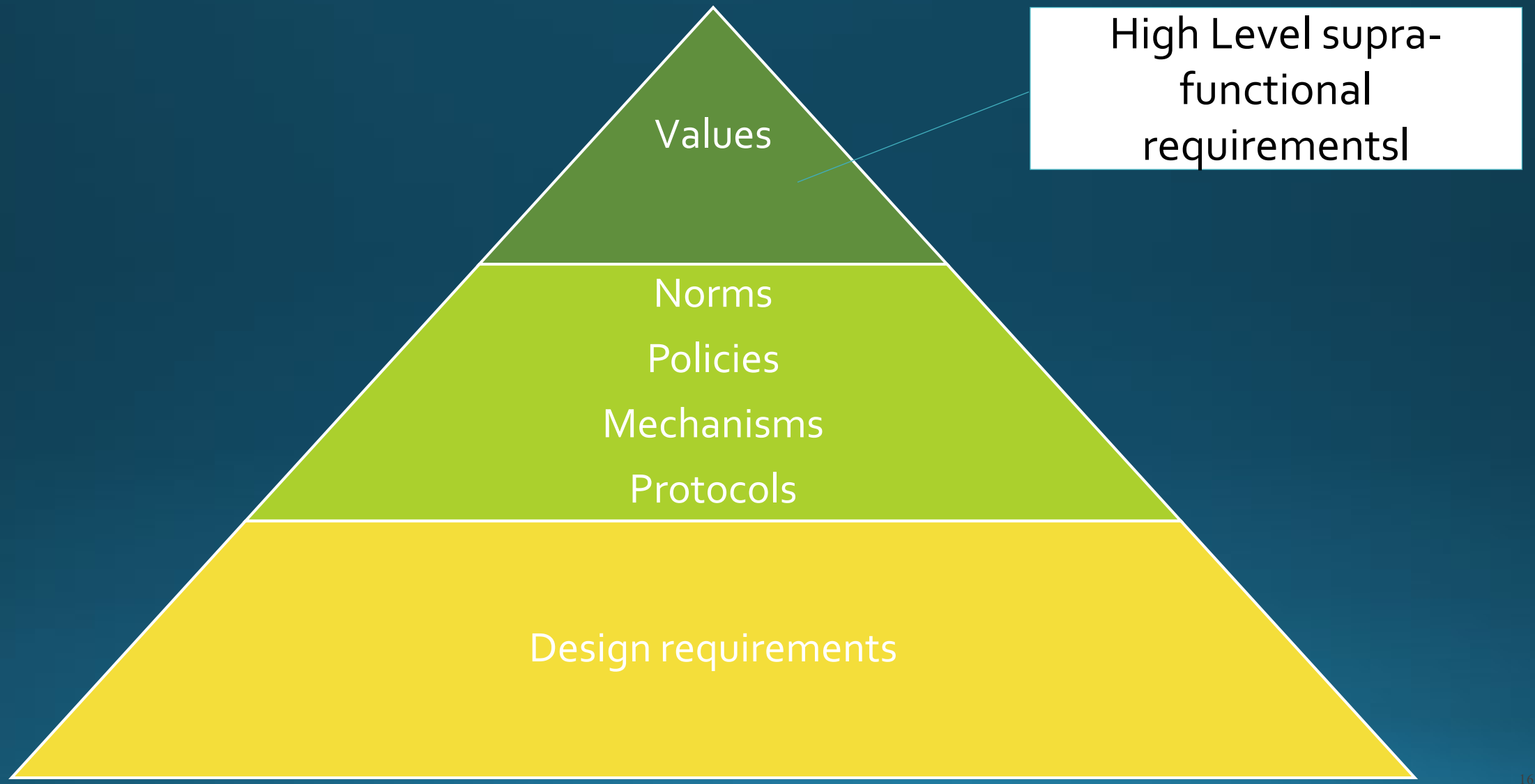
Key Problem
21st Century: Value Sensitive Design

Design for X

- Design for privacy
- Design for security
- Design for inclusion
- Design for sustainability
- Design for democracy
- Design for safety
- Design for transparency
- Design for accountability
- Design for responsibility



Values hierarchy



Example of values hierarchy

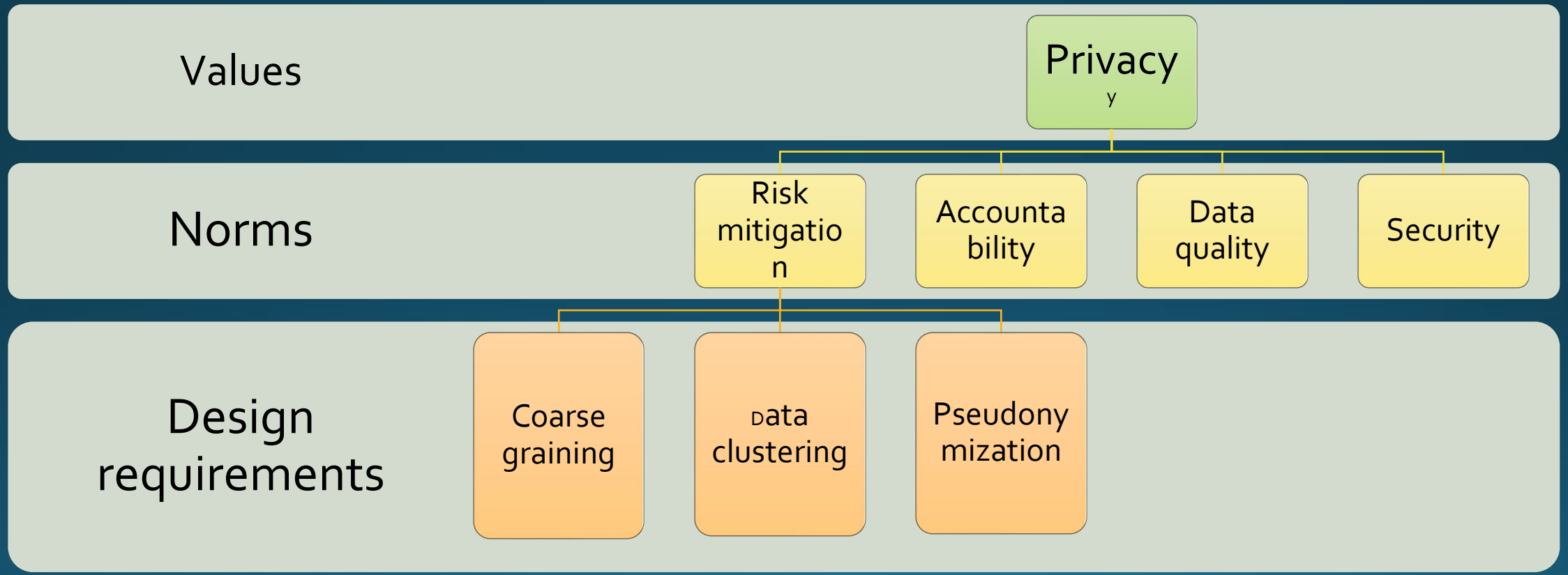
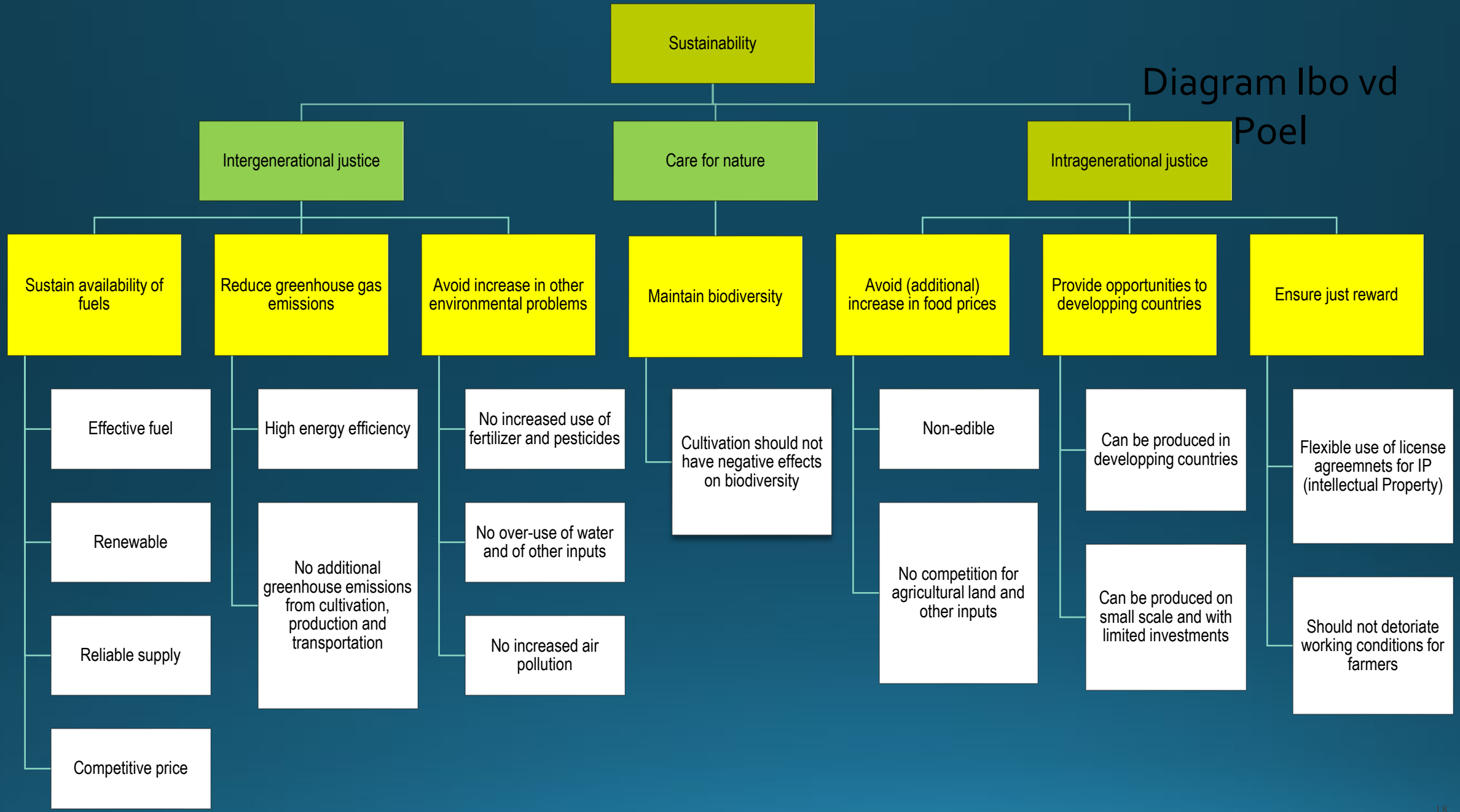


Diagram Ibo vd Poel

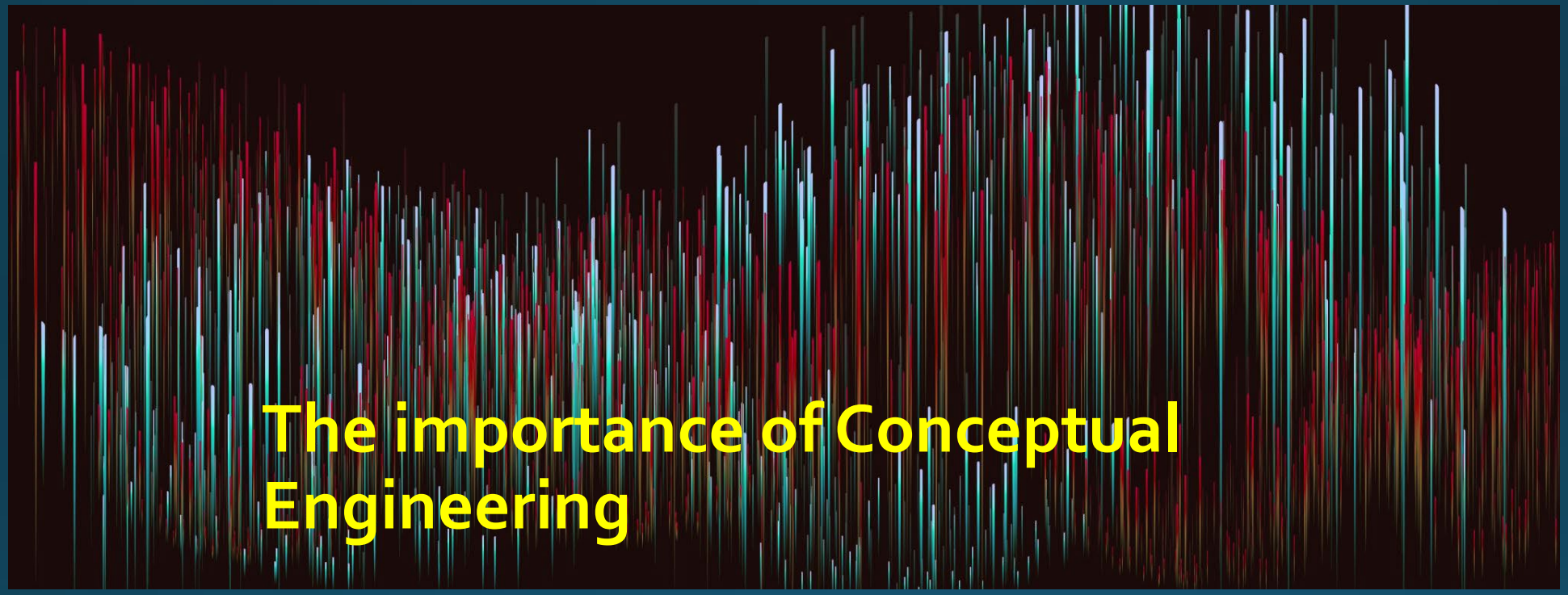


Algorithmic Fairness

"21 Definitions of Algorithmic Fairness"

- There are more than 30 different mathematical definitions of fairness in the academic literature.
- There isn't a one, true definition of fairness.
- These definitions can be grouped together into three families:
  - Anti-Classification
  - Classification Parity
  - Calibration

Arvind Narayanan


The importance of Conceptual Engineering

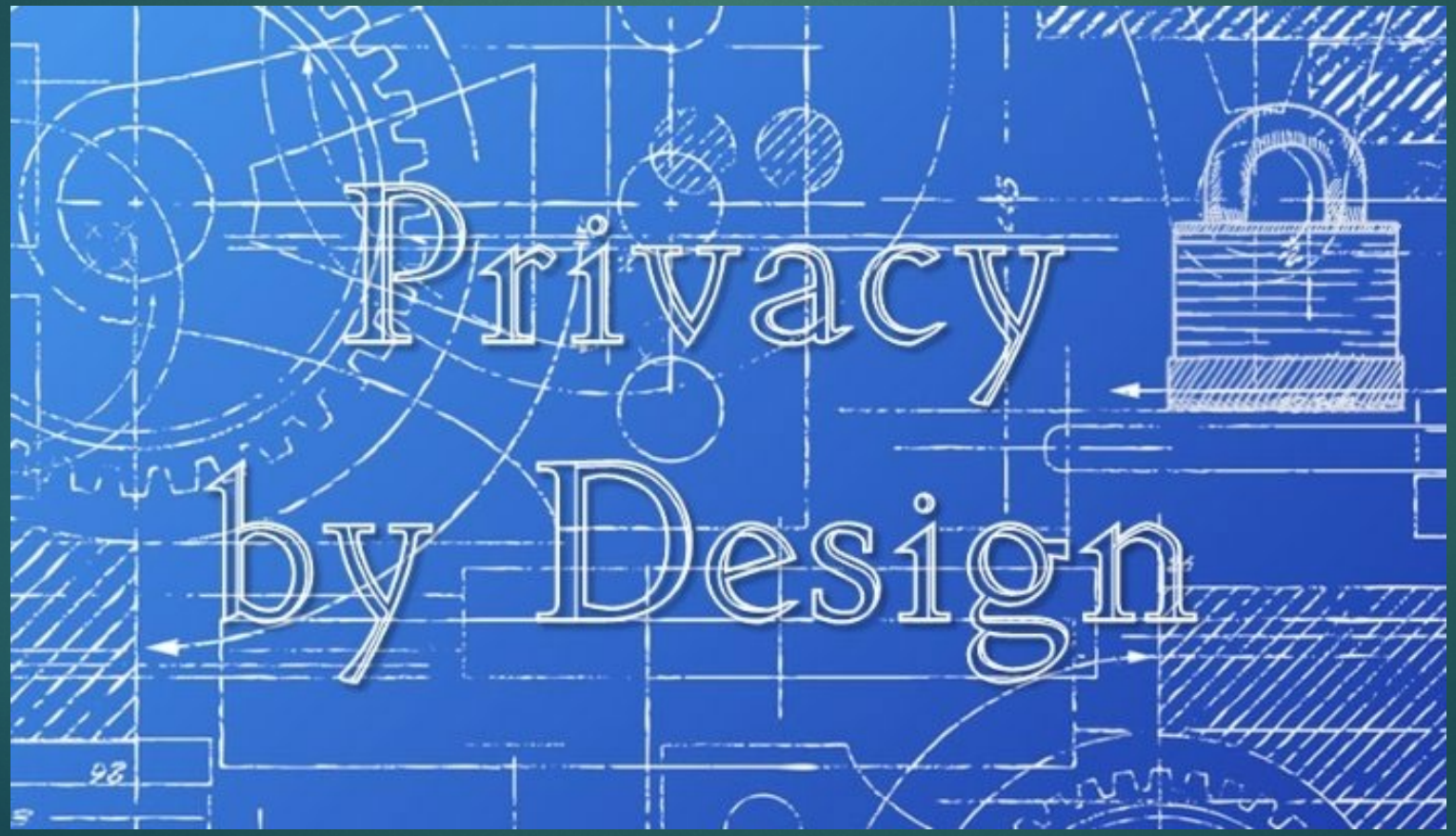
Ethics by Design

1
- Conceptual Engineering
- Moral Values: non-functional requirements

2
- Requirements Engineering
- Functional decomposition

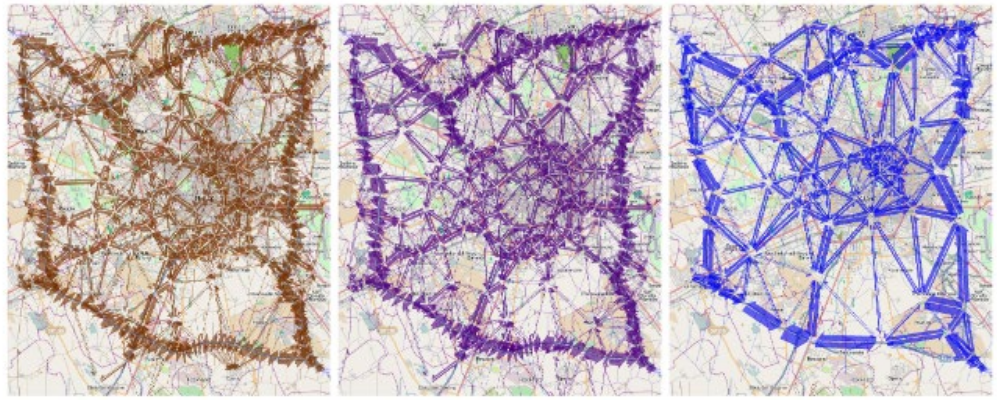
3
- Engineering Design
- Protypes and validation

Innovations in Privacy by design



Coarse graining through clustering

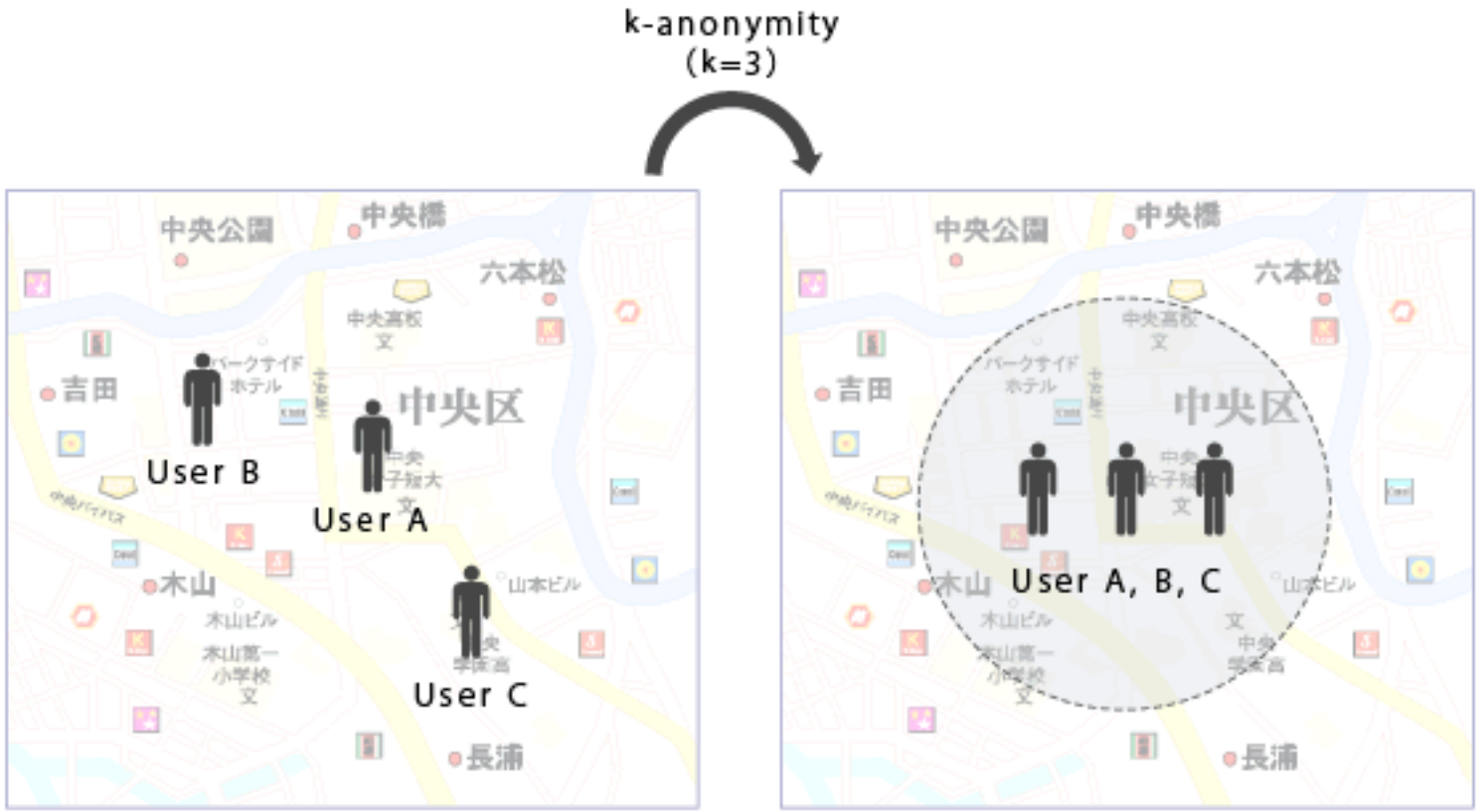
Monreale et al. clustering spatio-temporal trajectories: so that it remains possible to gauge the traffic in a city (utility), while making it impossible to e.g. stalk an individual citizen (privacy).



(a) (b) (c)

Figure 3: The results of progressive generalization after 11 (a), 12 (b) and 36 (c) iteration steps.

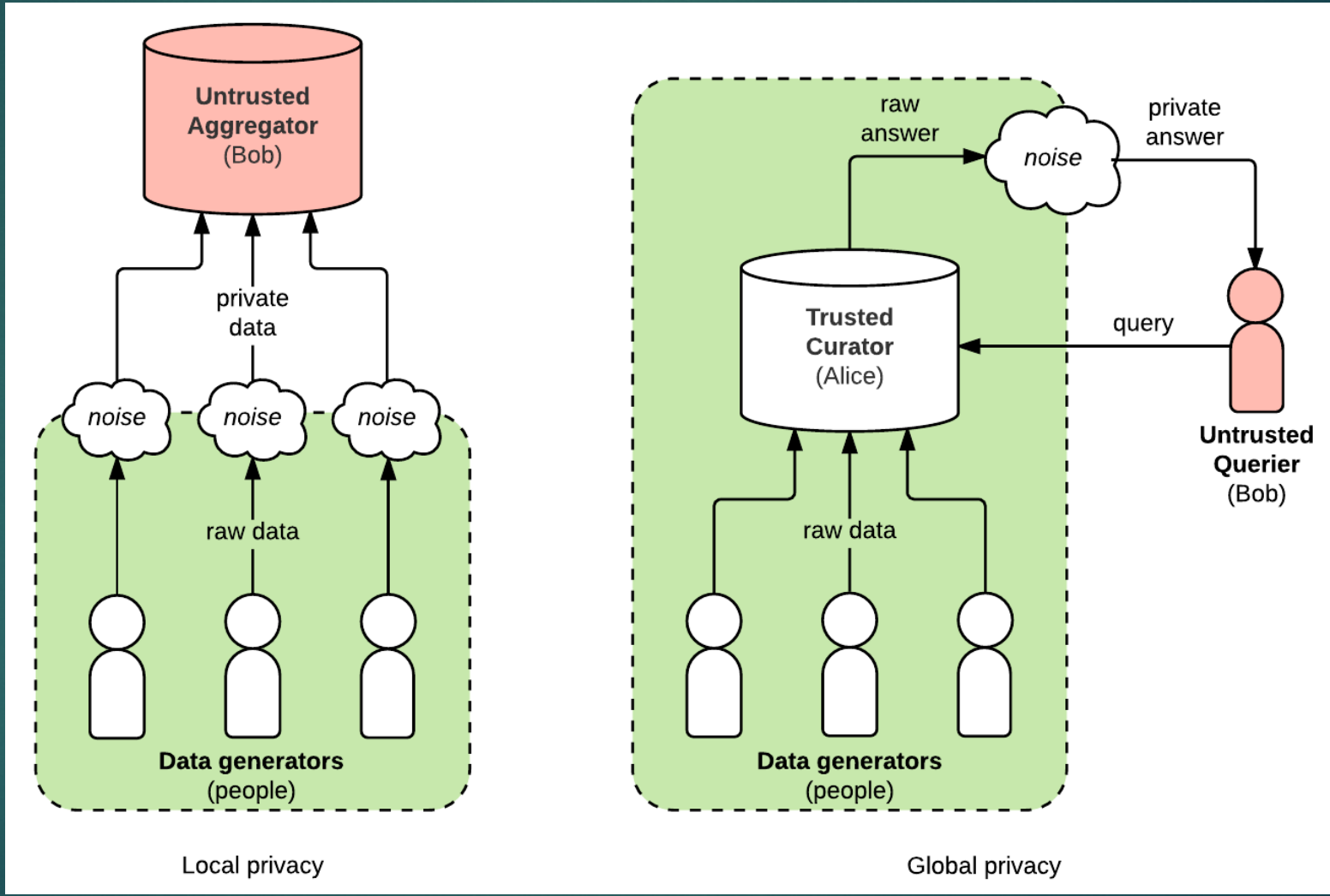
K-anonymity



Can identify the user's detailed location from latitude and longitude.

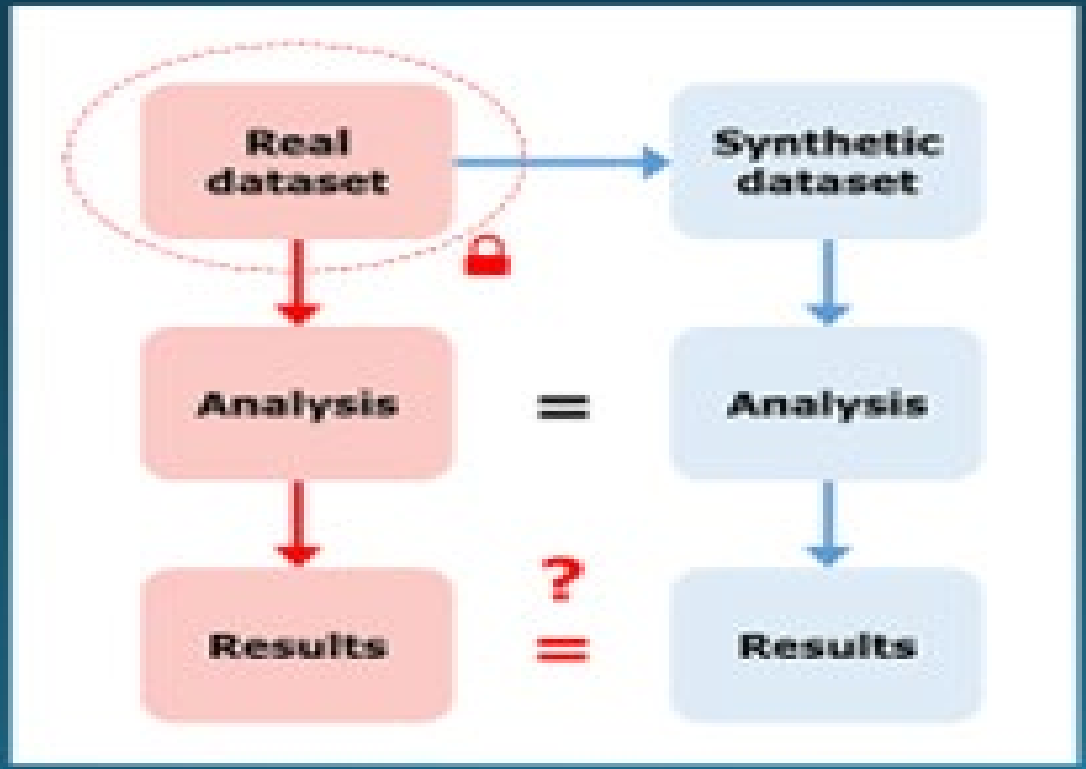
When the location information is blurred, it becomes impossible to tell who is where in the circle.

Differential Privacy




Homomorphic Encryption for Privacy-Preserving Machine Learning
AN INDUSTRY TALK WITH PROFESSIONALS
CSSS
computer science student society
ubccsss.org

Synthetic Healthcare Data



PUFFLE: Balancing Privacy, Utility, and Fairness in Federated Learning

Luca Corbucci [a,*], **Mikko Heikkilä** [b], **David Solans Noguero** [b], **Anna Monreale** [a] **and Nicolas Kourtellis** [b]

[a]University of Pisa
[b]Telefónica Research

Abstract. Training and deploying Machine Learning models that simultaneously adhere to principles of fairness and privacy while ensuring good utility poses a significant challenge. The interplay between these three factors of trustworthiness is frequently underestimated and remains insufficiently explored. Consequently, many efforts focus on ensuring only two of these factors, neglecting one in the process. The decentralization of the datasets and the variations in distributions among the clients exacerbate the complexity of achieving this ethical trade-off in the context of Federated Learning (FL). For the first time in FL literature, we address these three factors of trustworthiness. We introduce PUFFLE, a high-level parameterised approach that can help in the exploration of the balance between utility, privacy, and fairness in FL scenarios. We prove that PUFFLE can be effective across diverse datasets, models, and data distributions, reducing the model unfairness up to 75%, with a maximum reduction in the utility of 17% in the worst scenario, while maintaining the context of FL, and present a methodology that enables informed decisions on the ethical trade-offs with the goal of training models that can strike a balance between these three dimensions.

Our contributions: We propose PUFFLE, a first-of-its-kind methodology for finding an optimal trade-off between privacy, utility and fairness while training an FL model. Our approach, inspired by the method in [53] for a centralised setting, aims to train a model that can satisfy specific fairness and privacy requirements through the active contribution of each client. In particular, the clients incorporate an additional regularization term into the local loss function during the model's training to reduce unfairness. Furthermore, to ensure the privacy of the model, each client employs Differential Privacy (DP) [19] by introducing controlled noise during the training process. We summarize our key contributions below:

- Responsible Innovation with AI

>>If you can change the world with AI innovations today, so that you can meet more of your responsibilities tomorrow, you have responsibility to innovate with AI today<

Delft Digital Ethics Centre

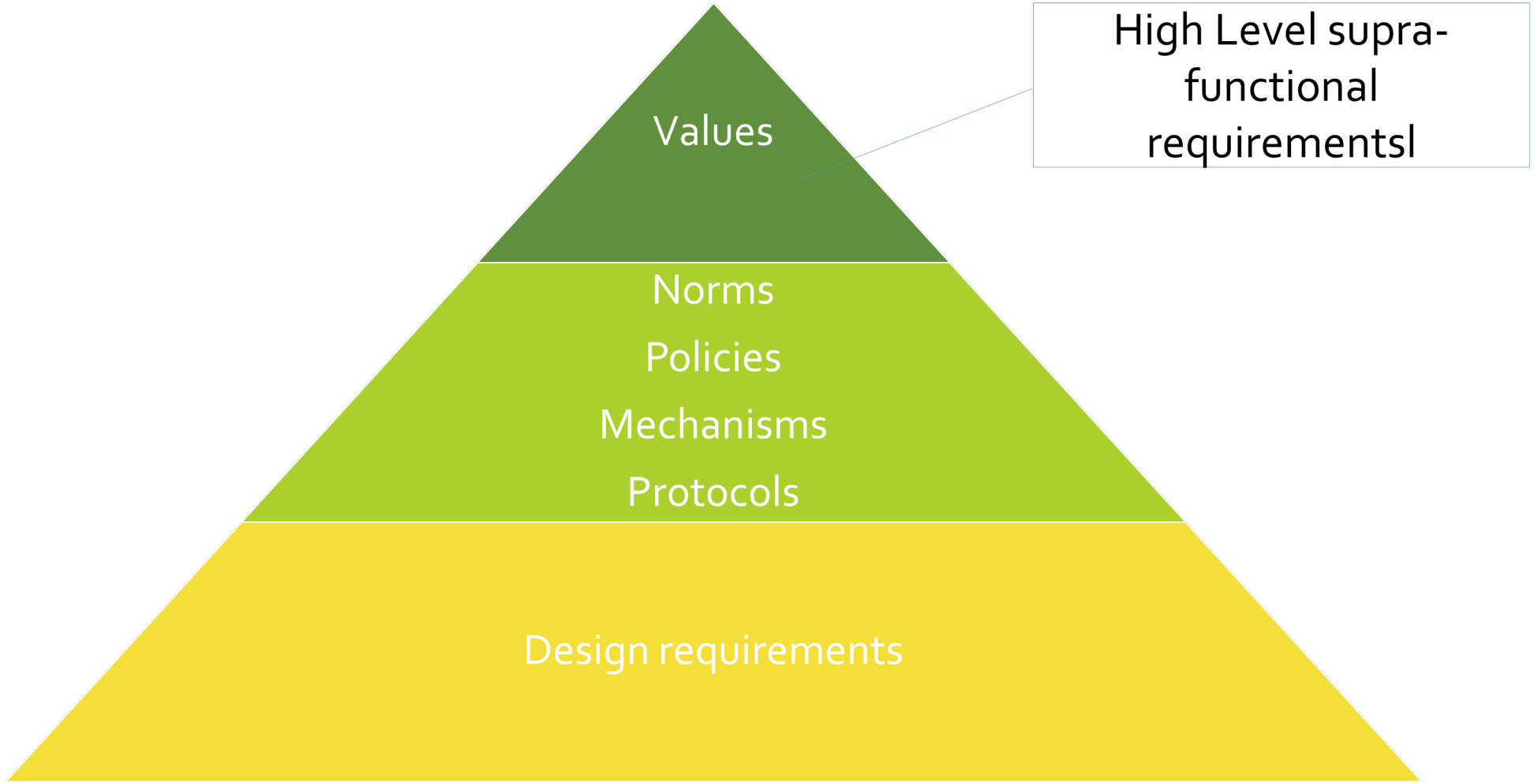
WHO Collaborating Centre on AI for health governance, including ethics

Where first-rate philosophy meets excellent engineering

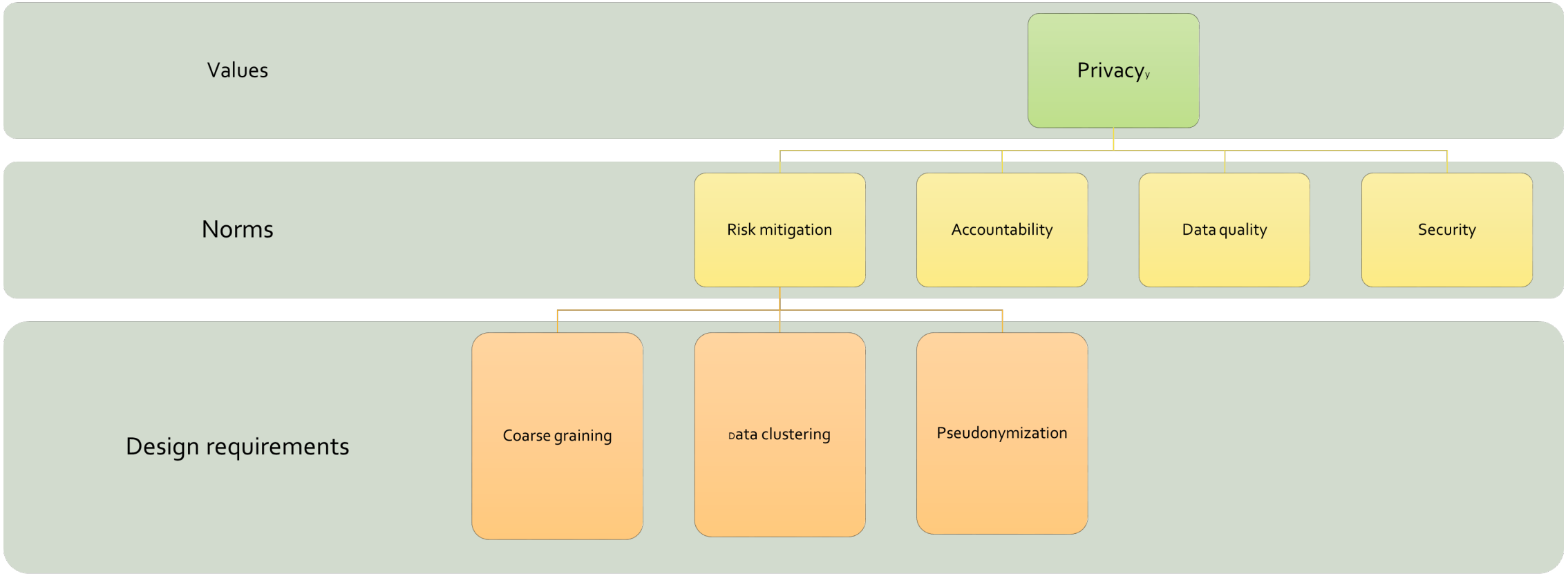
TU Delft

A growing research community of 30+ TU Delft researchers in the field of ethics and philosophy of digital technologies

Value hierarchy



Examples of value hierarchy



Value driven

Interdisciplinary

Collaborative

Aiming for innovations that address pressing societal and environmental challenges

Convergence Centre for **Responsible AI in Health Care**

REAiHL

Responsible and Ethical AI for Healthcare Lab

Design for Values

Research community

Ethics and philosophy of digital technologies

Conceptual Engineering

Dealing with Value Conflicts by Design

Values as (non) functional requirements

Conceptual understanding

of values central to the responsible development and use

Responsible Innovation

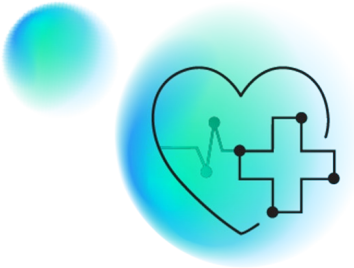
Applied projects

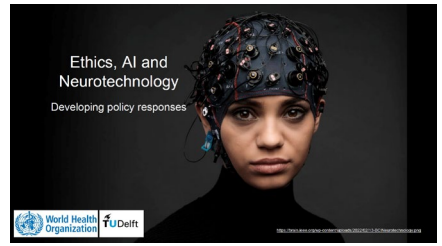
Translating values into actionable design requirements

Current successes together with our partners

REAiHL
Responsible and Ethical AI for Healthcare Lab

Dutch Adaptive Consortium for
Responsible Implementation of
Generative AI in Healthcare: Together





ICAI lab

- Bringing responsible AI to the clinic
- EMC, TU Delft, SAS

Indicate

- EU project, federated learning for ICU data
- EMC in the lead

RIGH:T consortium

- Clinical validation of GenAI – ambient listening
- Hospital lead consortium

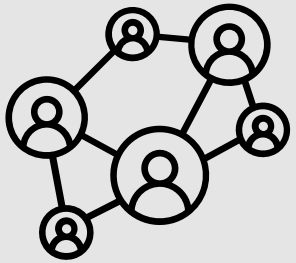
CRAIHC

- International Conference Responsible AI in Healthcare
- September 11 and 12, 2025
- Rotterdam

Neuro - AI

- Positioning paper with WHO
- In collaboration with EU

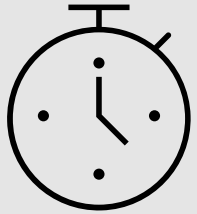
Our approach to Responsible AI for Healthcare



Collaborate

Actively set-up networks

Integrate ethics by design



Speed

Scale

Strategic autonomy

TU Delft

Thank you

We are looking forward to collaborating even more in the near future