WINTER SCHOOL
RESPONSIBLE MACHINE LEARNING

Umeå, 11-13 March 2025

LEMUR

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

WASP—HS

AEQUITAS

UMEÅ UNIVERSITY

AI Policy Lab
at Umeå University

Begränsad delning

Value interpretation – Exercise 1

- Kidney transplantation:
- Patients awaiting kidney transplantation are classified into 3 groups—low, medium, and high risk of kidney rejection.
- A large medical dataset is available with information about risk of rejection for patients.
- Traditionally, prioritization only takes the patient's age into account.
- You are hired to develop an AI application that improves kidney assignment decision

Adapted from: Sabine N van der Veer and others, Trading off accuracy and explainability in AI decision-making: findings from 2 citizens' juries, *Journal of the American Medical Informatics Association*, Volume 28, Issue 10, October 2021, Pages 2128–2138, https://doi.org/10.1093/jamia/ocab127

Begränsad delning

Question 0

- Question: Is AI suitable to develop a system to categorize patients awaiting kidney transplantation ?

- Options
  - A) Yes, AI systems guarantee most accurate and robust results
  - B) Yes, if AI system ensures that explanations are available
  - C) Yes, if AI system is as fair as possible, i.e. it does not take into consideration any personal characteristics on the decision
  - D) No. Decisions about life and dead must be done by experts.

Join at

slido.com

#4275 548

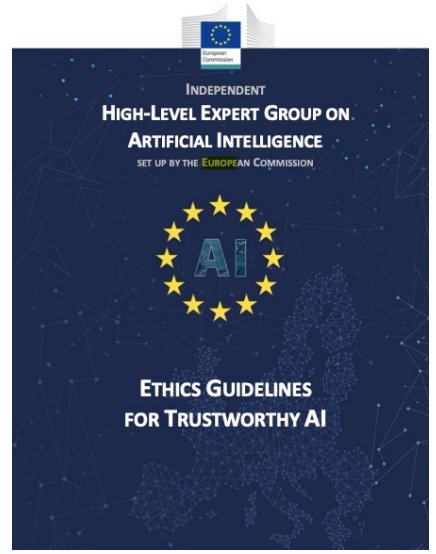
Begränsad delning

Values

- Value alignment
  - Which values
  - Who gets to decide
- System characteristics
  - Which methods
  - Which requirements?

Rank the requirements

- Accuracy
- Robustness
- Accountability
- Human Oversight
- Privacy
- Explainability
- Transparency
- Non-discrimination and fairness
- Environmental wellbeing



https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

Begränsad delning

Trade-offs

How to split your resources:

- How much (computational) cost is worth to include explanation?
- How much (computational) cost is worth 5% more accuracy?

What is fair?

- All patients are prioritized based only on age, treating everyone the same.
- Priority is based on risk level, giving high-risk patients better-matched organs.

Begränsad delning

Dilemmas – Homework

- Recruitment: Now you will be developing an AI system to screen applicants for a job, offering an interview based on their likelihood to become a “high-performing employee” using existing company data. This replaces the need for manual short-listing and telephone pre-interviews and provides a more just decision for both candidates and company

Adapted from: Sabine N van der Veer and others, Trading off accuracy and explainability in AI decision-making: findings from 2 citizens’ juries, *Journal of the American Medical Informatics Association*, Volume 28, Issue 10, October 2021, Pages 2128–2138, https://doi.org/10.1093/jamia/ocab127