

**Amann et al. (2020):  
Explainability for Artificial  
Intelligence in Healthcare:  
A Multidisciplinary Perspective**



# Explainability

Amann et al., p. 2:

- ▶ “In its essence, explainability can be understood as a characteristic of an AI-driven system allowing a person to reconstruct why a certain AI came up with the presented predictions.”

# Discussing the role of explainability in AI from various perspectives

- ▶ The technological perspective
- ▶ The legal perspective
- ▶ The medical perspective
- ▶ The patient perspective

Context:

AI-based Clinical Decision Support Systems (CDSS)



# Technological perspective

- ▶ Explainability („interpretability“) as an inherent characteristic of algorithms (e.g., linear or logistic regression)
- ▶ Methods that approximate explainability (e.g. LIME, SHAP)
- ▶ Trade-off between performance and explainability
- ▶ Explainability helps developers sanity check their models

# Legal perspective

- ▶ Is explainability in AI legally required? If so, to what extent?
- ▶ Informed consent
  - Precondition for processing personal health data
  - Requires adequate information and understanding
- ▶ Certification and approval as medical devices
  - Role of explainability is currently unclear
- ▶ Liability
  - Need for disclosure of the use of an AI-based CDSS?
  - Legal and litigation questions, depending on whether or not a physician adheres to the AI recommendation

# Medical perspective

- ▶ Need to reflect on how AI-based clinical decision support is different from established diagnostic tools
- ▶ Need for clinical validation
  - Accuracy
  - Bias leading to systematic errors
- ▶ “Explainability enables the resolution of disagreement between an AI system and human experts“ (p. 5)

# Patient perspective

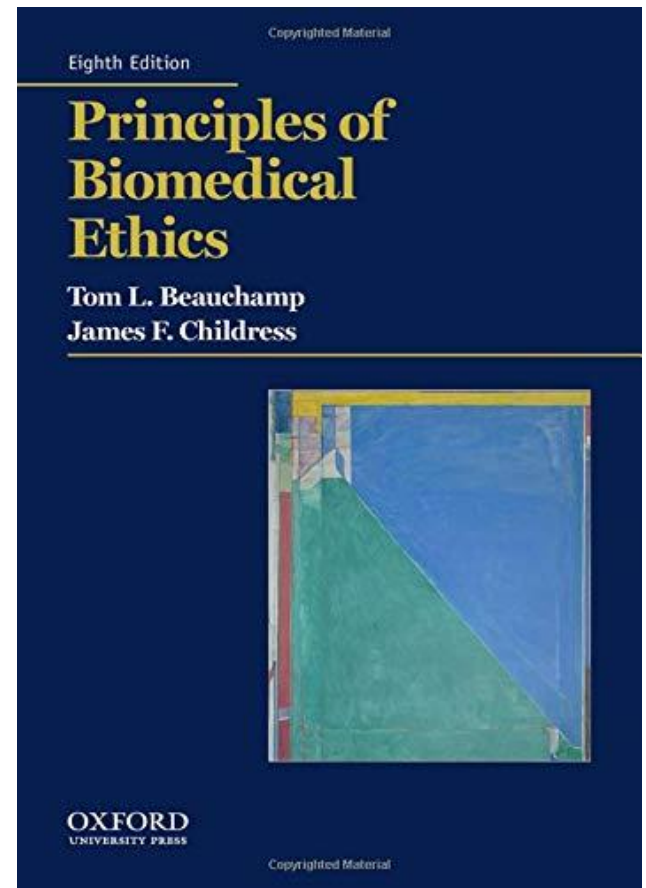
- ▶ Black box AI conflicts with the core ideals of patient-centered medicine
- ▶ Clinicians need to understand the relevant aspects in order to explain them to patients
- ▶ Explainability facilitates conversations between medical doctors and patients

# Tom L. Beauchamp & James F. Childress: Principles of Biomedical Ethics

- ▶ Respect for Human Autonomy
- ▶ Beneficence
- ▶ Non-Maleficence
- ▶ Justice

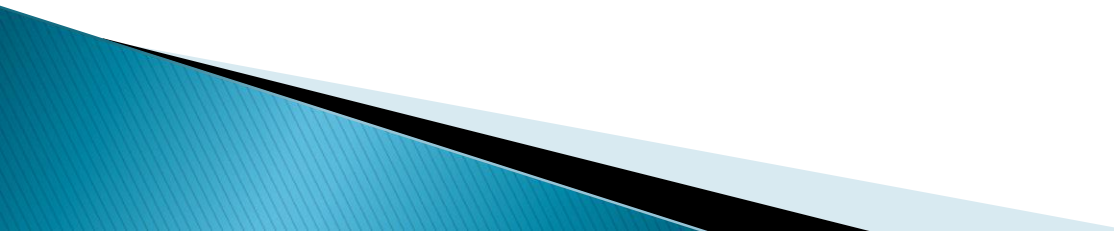


“Based on our analysis, we consider explainability a necessary requirement to address these challenges in a sustainable manner that is compatible with professional norms and values.”  
(Amann et al., p. 8)

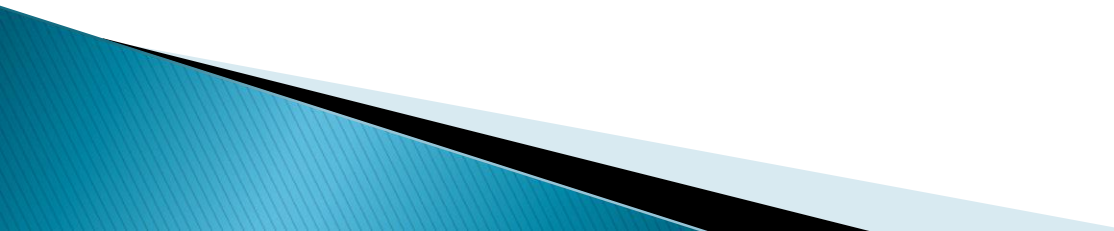




# Questions for discussion

- ▶ How is AI-based clinical decision support different from established diagnostic tools?
  - ▶ Who benefits from explainable AI?
- 

# Questions for discussion

- ▶ What type of explanation is needed? How much explanation is needed?
  - ▶ How important is explainability, compared to accuracy?
- 

# Questions for discussion

- ▶ Should medical doctors and healthcare professionals be required to let patients know when they use an AI-based clinical decision support system (CDSS)?
- ▶ How may the use of AI-based CDSSs influence patient autonomy and informed consent?

# Questions for discussion

- ▶ How may the use of AI-based CDSSs influence the autonomy of healthcare professionals?
  - ▶ Is there a risk that healthcare professionals tend to follow the AI-based CDSS output in order to avoid legal and medical repercussions?
- 