



# Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare

Jonathan Spsychalski, Evan Bestic, Michael Greer,  
and Connor McDonald



# The Increasing Prevalence of AI

- Applications across various fields
  - Econometrics, Biometry, E-Commerce, Automotives...
- Widely used in Healthcare
  - Clinical Decision Support (CDS)
  - Medical Imaging Analysis
  - Analysis of Pathology and Radiology Reports
  - Emerging Applications of LLMs



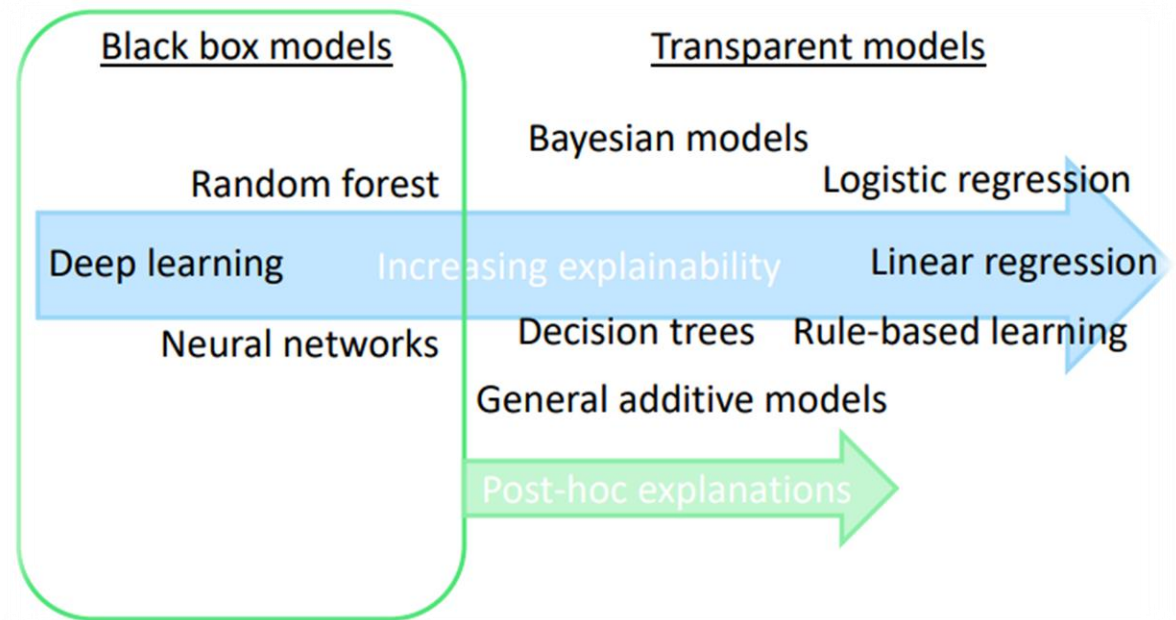
# Importance of XAI in Healthcare

- AI in healthcare presents various challenges
  - Bias
  - Privacy and Security
  - Trust and Accuracy
- Explainable AI (XAI) can mitigate these issues
  - XAI can reveal issues of bias and privacy in training data
  - Transparent models will be more trusted by users



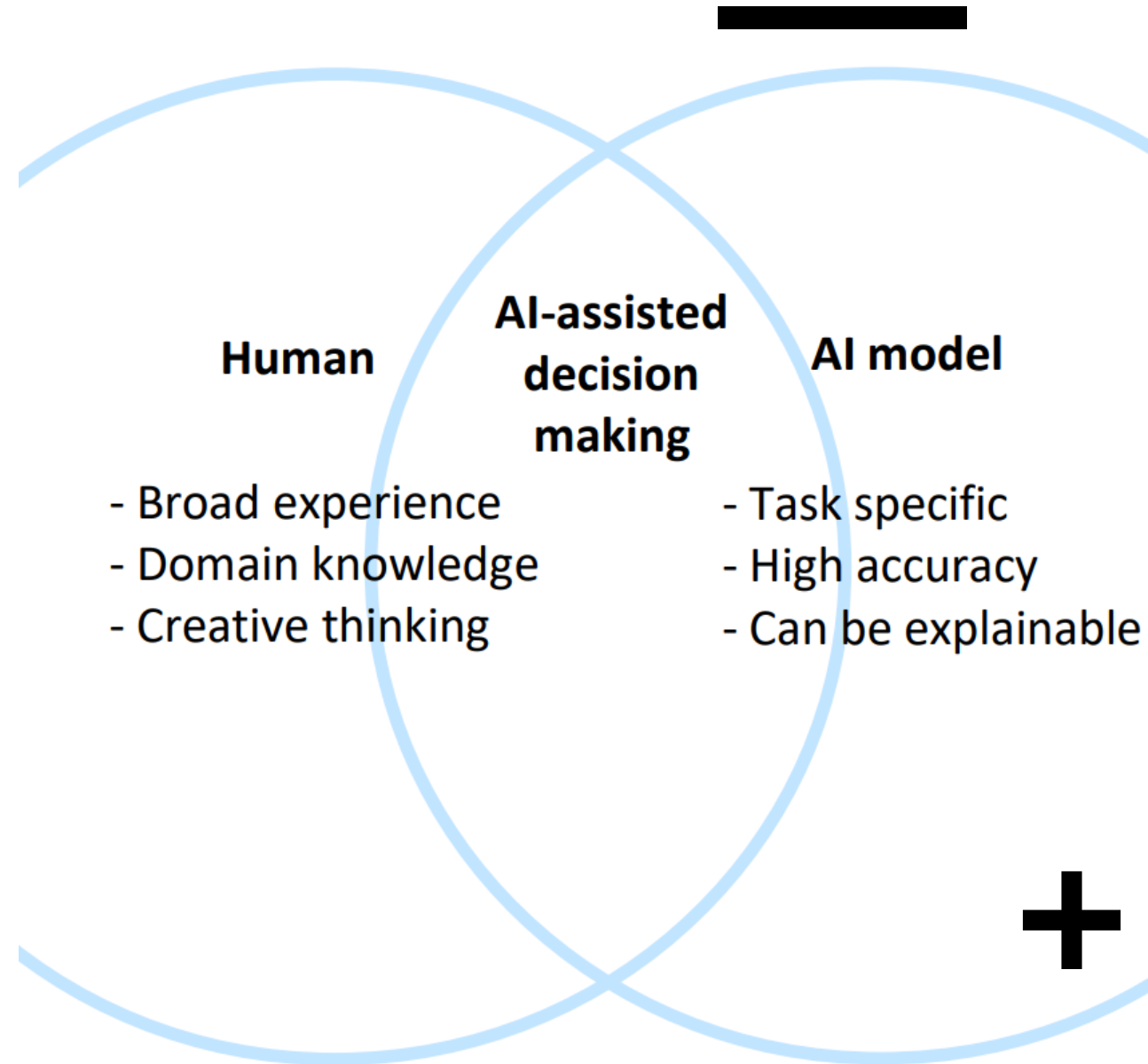
# XAI Taxonomy

- **Ante-Hoc:** Inherently explainable models
  - Simple models (e.g., decision trees)
  - No additional explanation needed
- **Post-Hoc:** “Black-box” models with layer of added explanation
  - Complex models (e.g., neural nets)
  - May require diverse means of explanation



# Human-XAI Collaboration

- Humans must not be “replaced” by AI at any stage of diagnosis or treatment
- Rather, clinicians and patients should work in tandem with AI to get the most value from it
- XAI enables this collaboration
  - “Explainability allows developers to identify shortcomings in a system and allows clinicians to be confident in the decisions they make with the support of AI”



# Promising XAI Paradigms for Healthcare

- Scientific XAI

- Focuses on producing a bona fide scientific explanation, rather than simply describing “how” a model arrived at a medical decision

- Granular Computing (GrC)

- Information is divided and processed in “granules”. This can help connect abstract concepts and decisions to concrete data

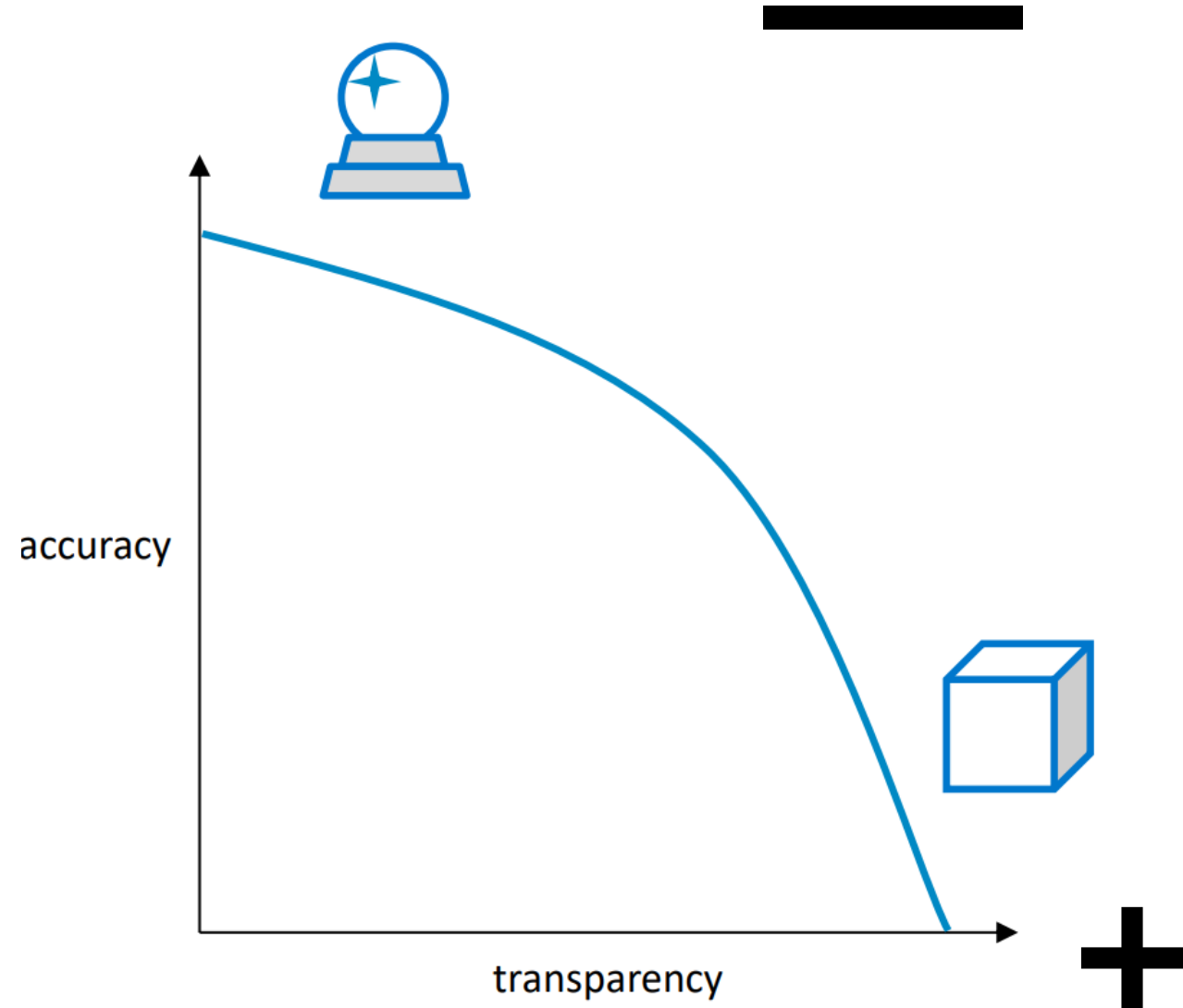
- Fuzzy Modeling (FM)

- Represents a model’s decision-making process in a way that is more intuitive and interpretable; useful complement to GrC



# Challenges for XAI in Healthcare

- Privacy and security
- Accuracy-explainability trade-off
- “Overtrust” in AI decisions
- Assessing Explainability
- Increasing complexity of models



# Discussion

- Do we feel comfortable allowing AI to make predictions and be used in healthcare settings, such as for analyzing scans?
- Will AI remain in a purely supporting role for clinicians, or will it eventually make critical healthcare decisions without human involvement?
- Would you trust a usually accurate but opaque AI to make decisions about your healthcare? Or is explainability a necessity in healthcare?

