

Explainability for Artificial Intelligence in Healthcare

Lifang He, Ph.D.

Dept. of Computer Science & Engineering

September 15, 2023



Outline

- Introduction
- Explainable AI for Healthcare
 - What is Explainability in AI?
 - Why is Explainability Important in Healthcare?
 - Challenges in Explainability
 - Methods for Achieving Explainability
- Future Directions

Introduction

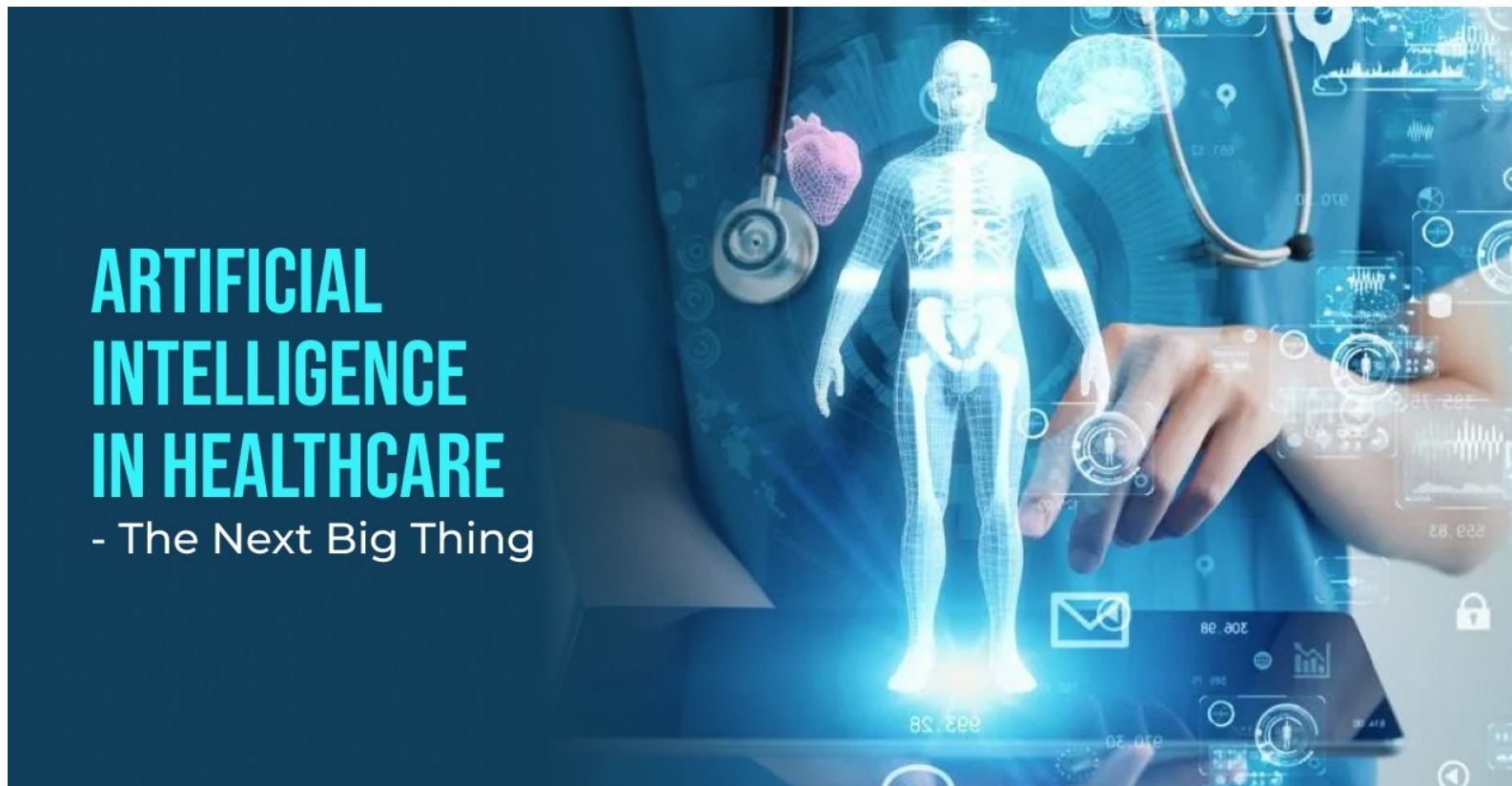
Importance of Healthcare

- Quality of Life
- Life Expectancy
- Social and Economic Impact



Role of AI in Healthcare

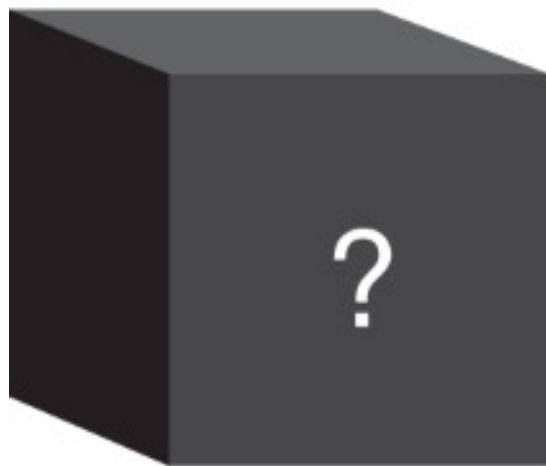
- Diagnosis and Treatment
- Personalized Medicine
- Drug Discovery
- Operations and Logistics



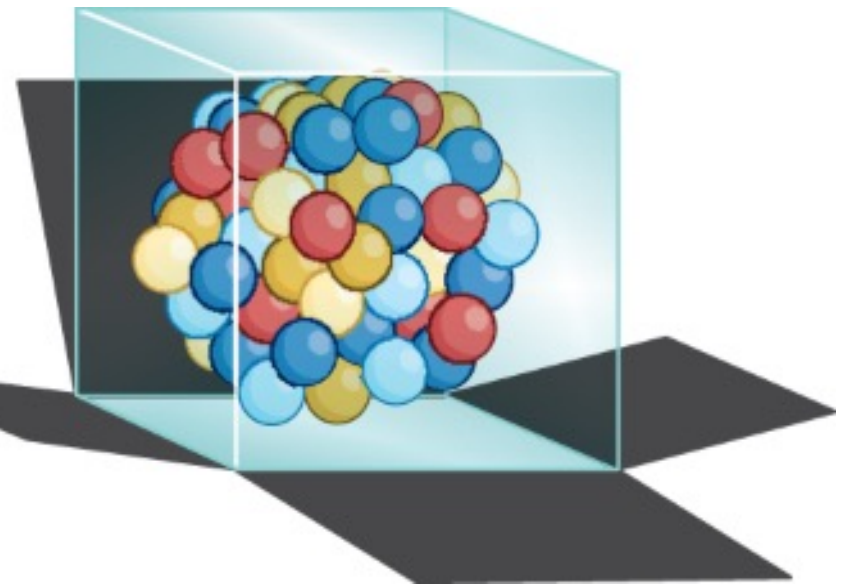
Explainable AI for Healthcare

What is Explainability?

- Explainability in the context of AI refers to the ability of a model to provide understandable and interpretable outputs or decisions.

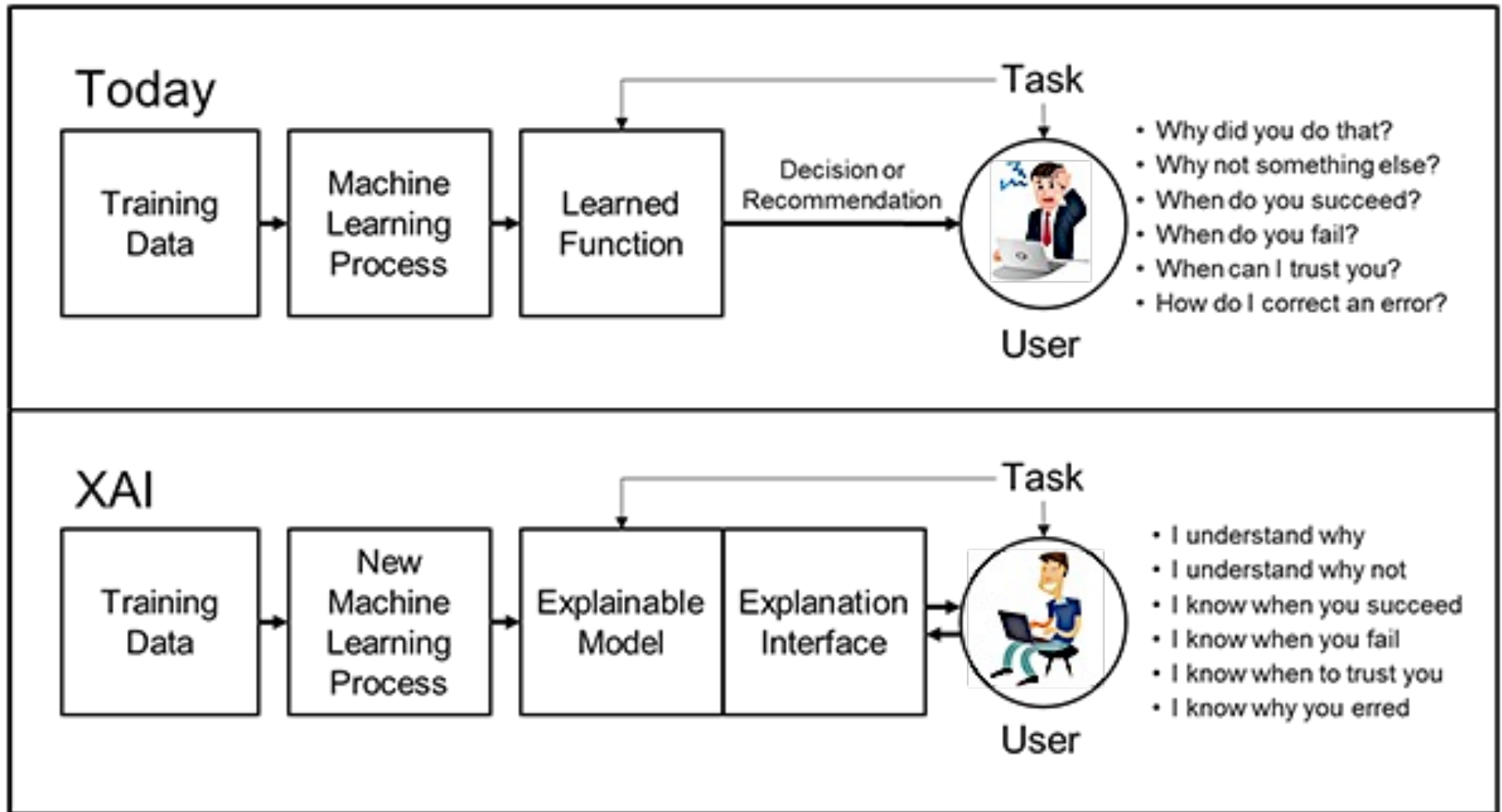


Black Box AI



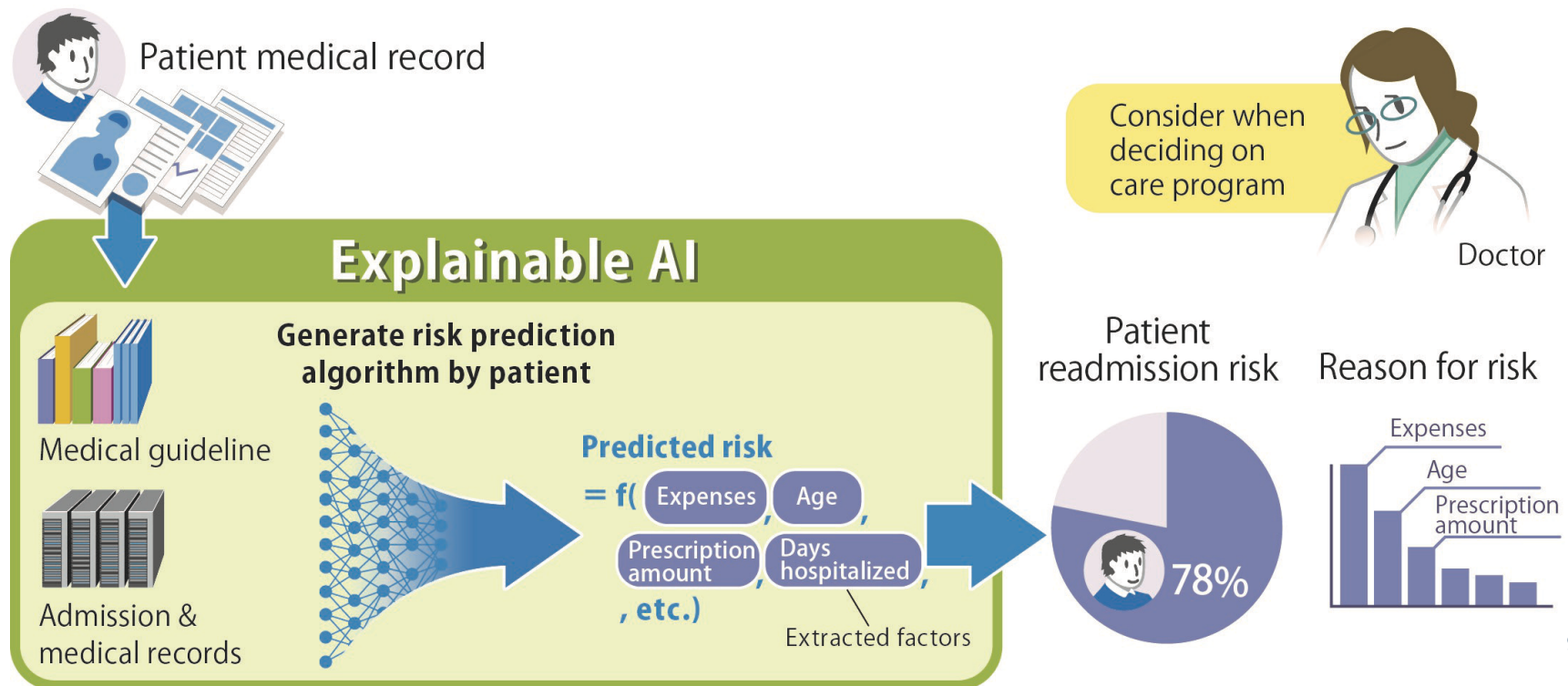
Explainable AI (XAI)

AI vs. Explainable AI



Why is Explainability Important in Healthcare?

- Transparency
- Trust
- Accountability
- Medical Ethics and Legal Compliance
- Reducing Impact of Model Biasing



Challenges in Explainability

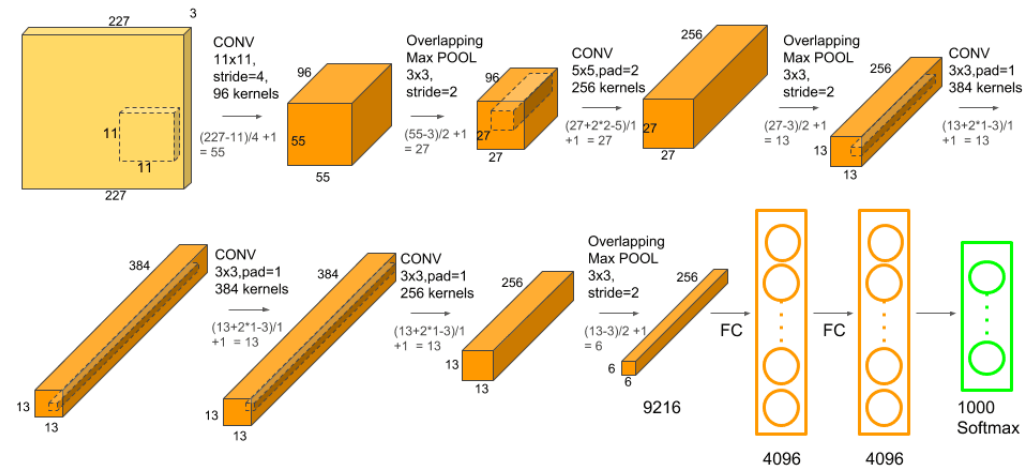
- Trade-off between Performance and Explainability
- Complexity of Models

$$Y = 100 * \text{age} + 10 * \text{income} + 200$$

$$Y = 100 * 29 + 10 * 3000 + 200 = \$33,100$$

Linear Regression

(3 parameters)

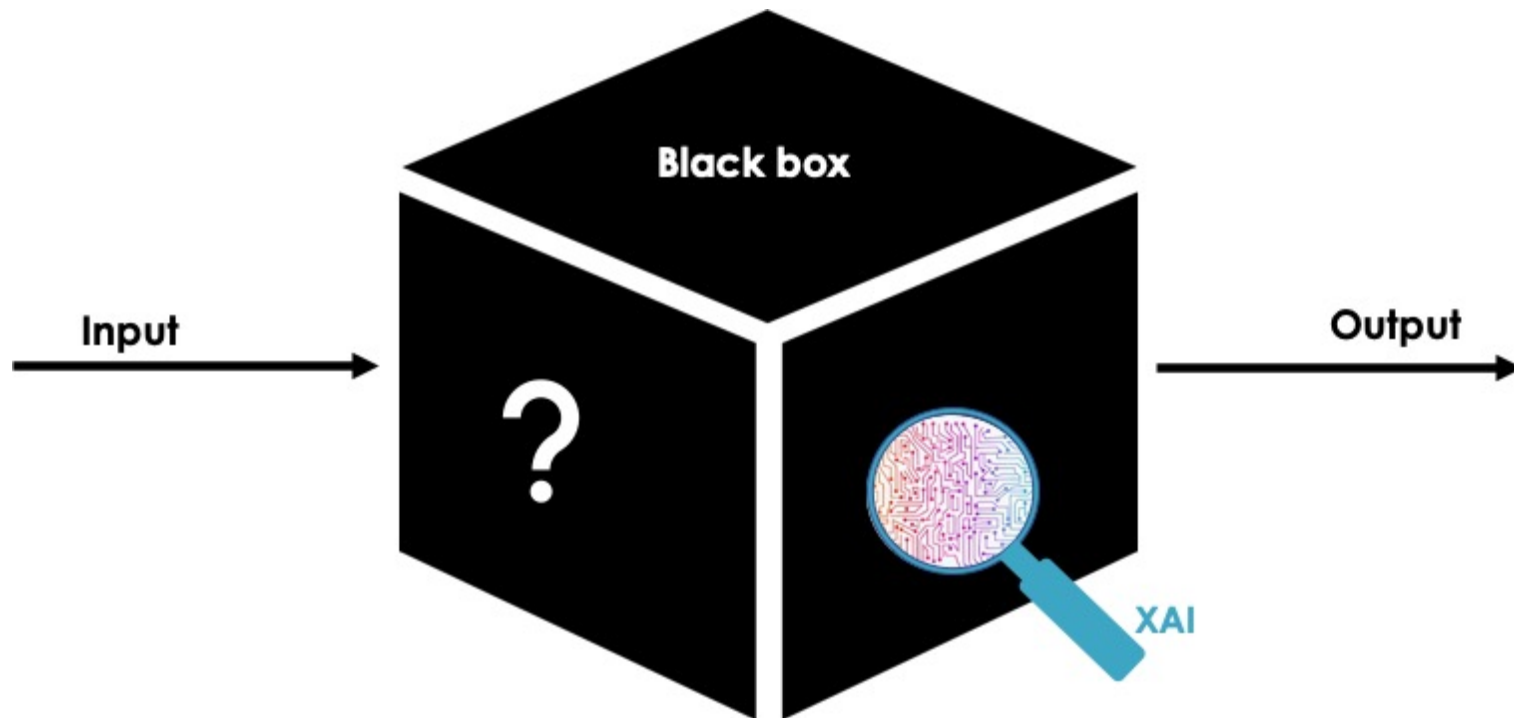


AlexNet

(62.3 million parameters)

Methods for Achieving Explainability

- Post-hoc Analysis
- Interpretable Models
- Visual Tools

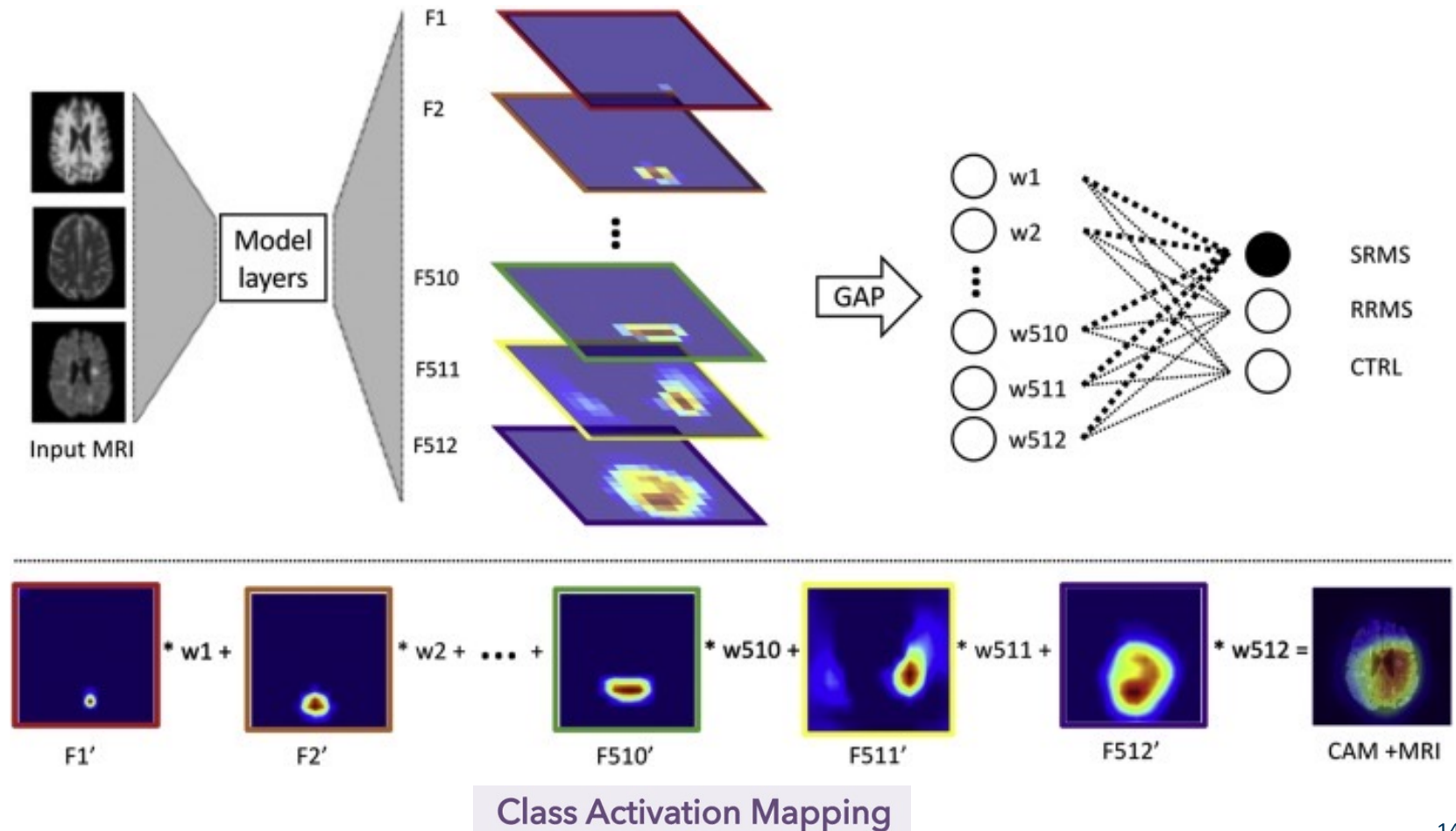


Examples of Explainable AI in Healthcare

- Case Study 1 - Post-hoc Analysis
- Case Study 2 - Interpretable Models
- Case Study 3 - Visual Tools

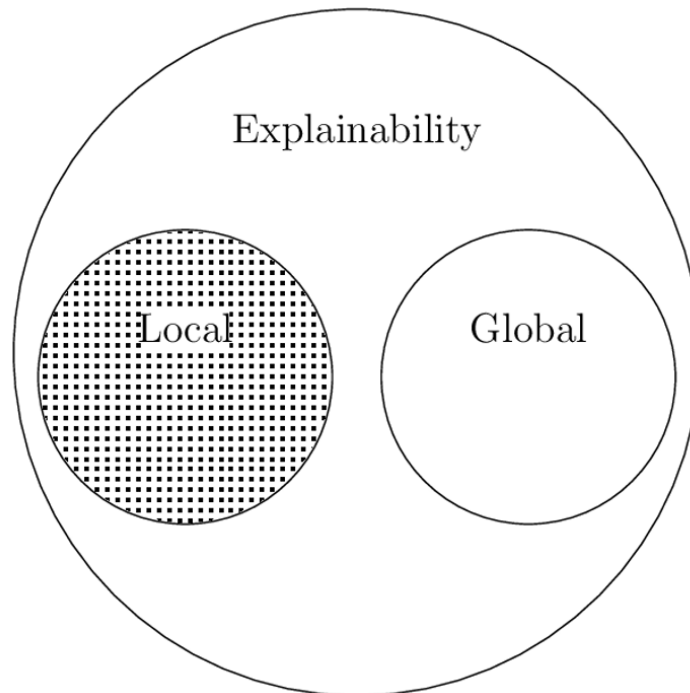
Case Study 1 - Post-hoc Analysis

- Gradient-weighted Class Activation Mapping (Grad-CAM)



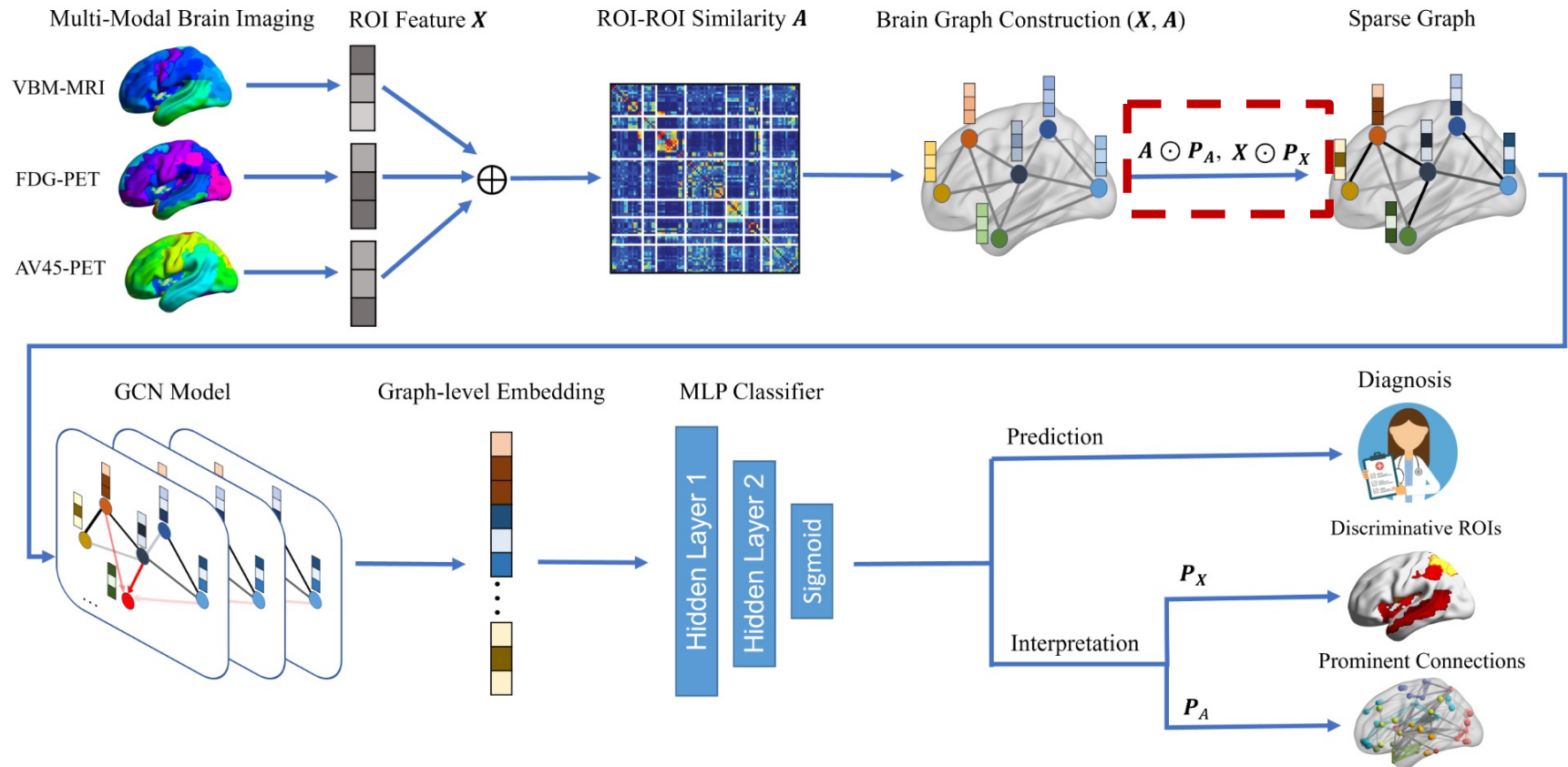
Case Study 1 - Post-hoc Analysis (Cont'd)

- **Local Explainability (e.g., LIME):** Understanding a specific decision made by the model.
- **Global Explainability (e.g., SHAP):** Understanding the entire model and how it makes decisions.



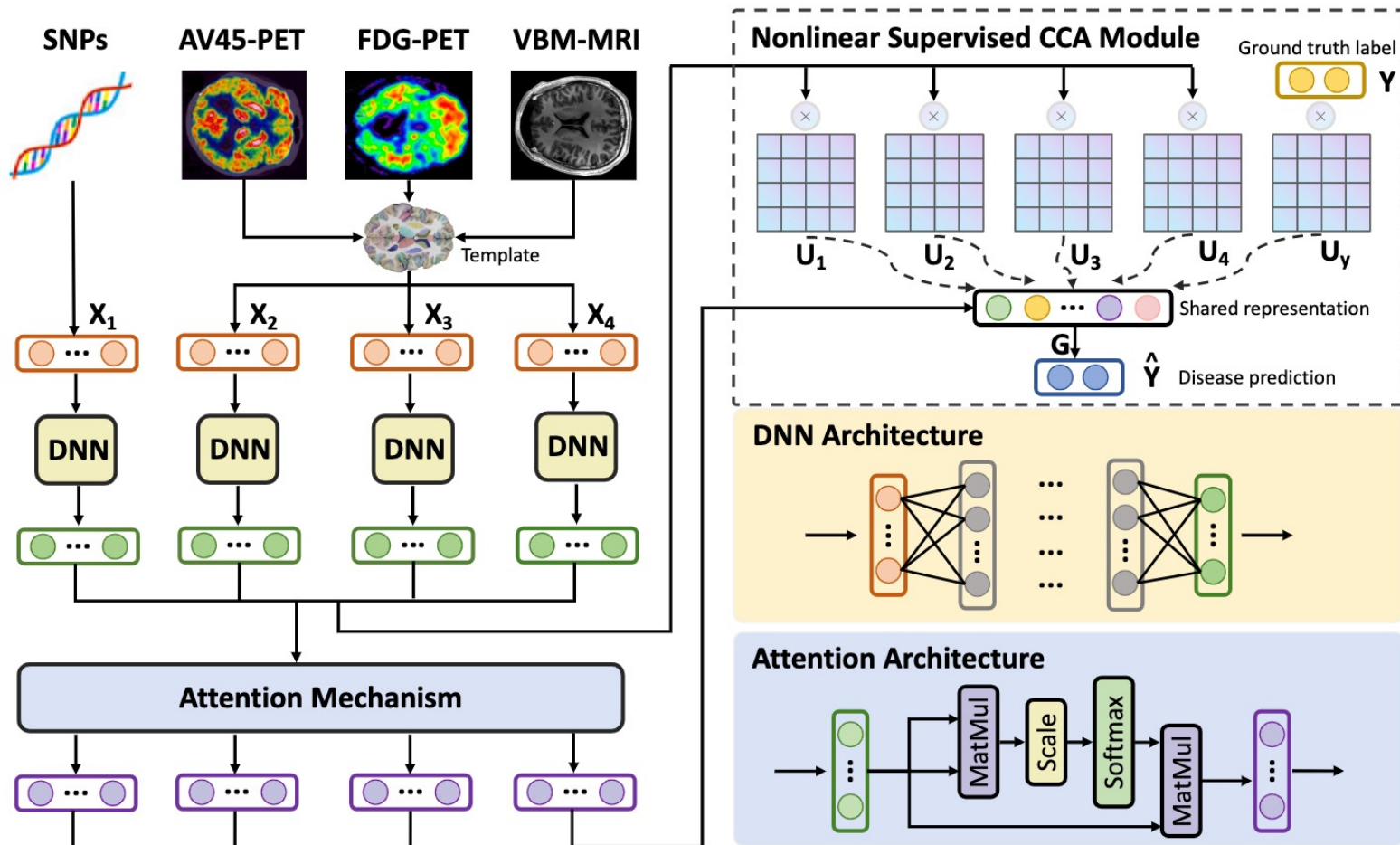
Case Study 2 - Interpretable Models

- Sparsity for Explanation



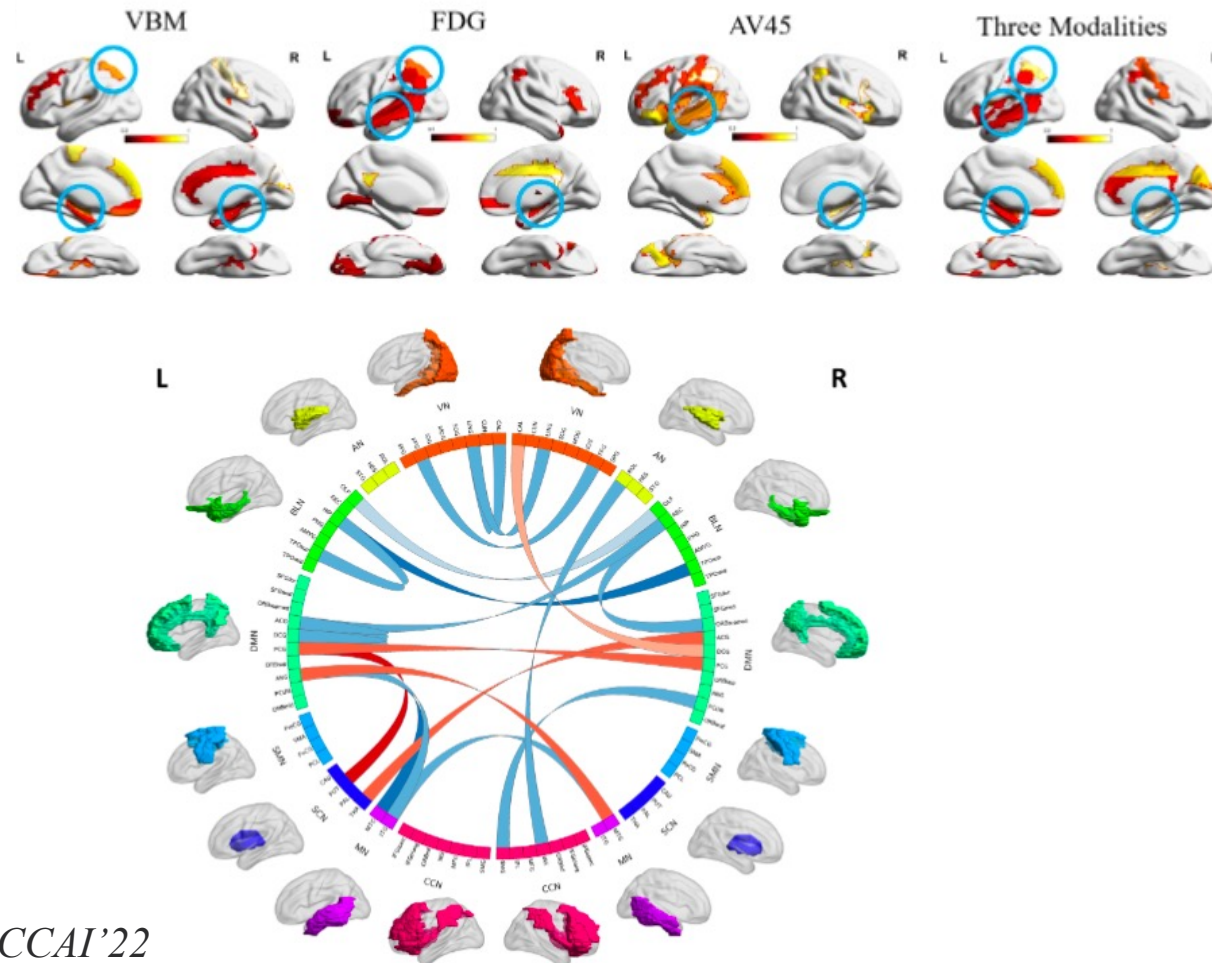
Case Study 2 - Interpretable Models (Cont'd)

- Attention Mechanisms for Explanation



Case Study 3 - Visual Tools

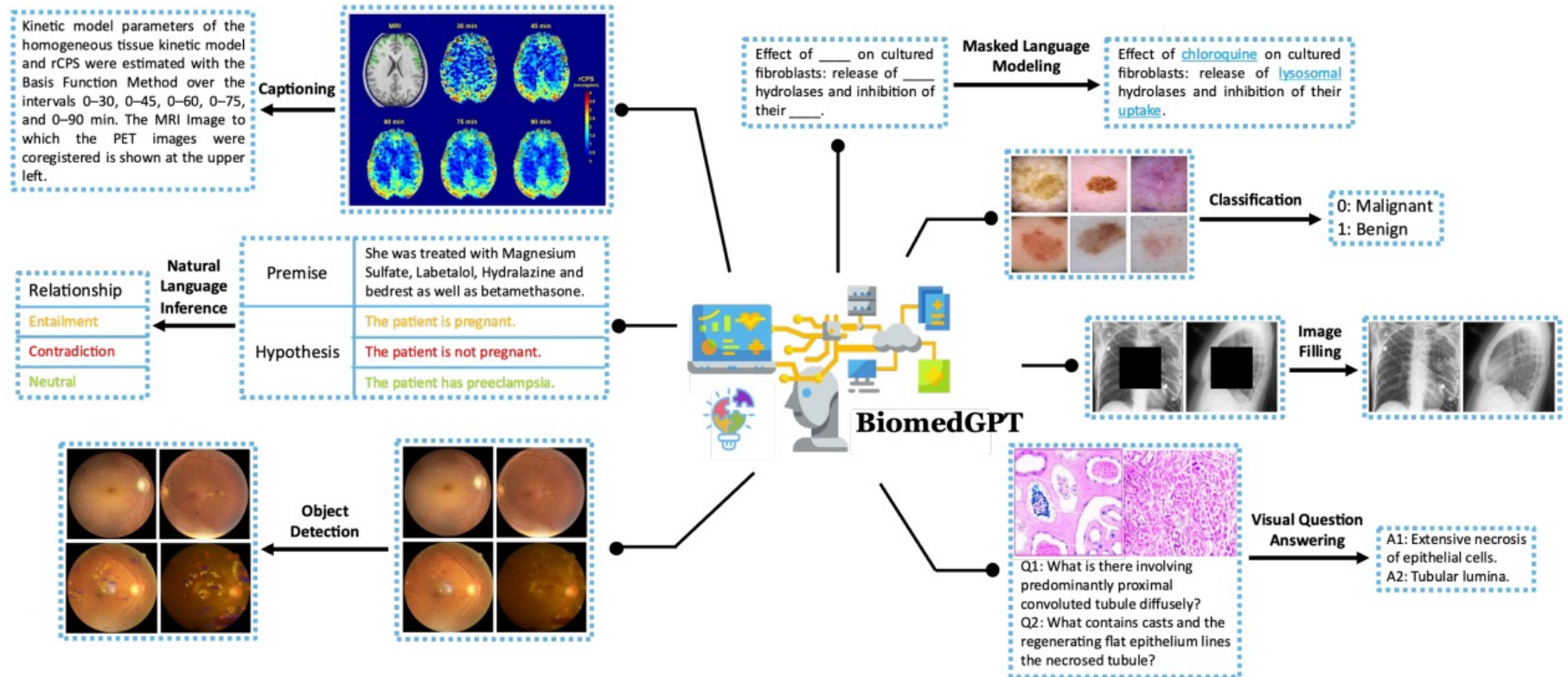
- Region and Connectivity Visualization for Brain Network



Future Directions

Future Trend in Healthcare

Harnessing Medical Data Variety and Other Technologies



Research and Development

- **Real-world Validation:** instead of controlled environments.
- **Explainability Metrics:** standardized metrics to evaluate XAI.
- **Human-AI Collaboration:** how AI can augment human decision-making in healthcare, rather than replace it.
- **Ethical AI:** creating AI models that are not just accurate but also ethical, unbiased, and just.
- **Interdisciplinary Research:** collaboration between healthcare professionals and AI researchers.

Policy Considerations

- **Regulatory Frameworks:** developing robust regulatory frameworks that can adapt to AI.
- **Data Privacy:** policies that protect patient data and ensure that it is used responsibly.
- **Transparency Standards:** policies that disclose how the AI models work.

Q&A

Lifang He, Ph.D.

lih319@lehigh.edu

*Thank
you*

