Clinical User-Centered Explainable AI for Medical Image Analysis
A Critical Technical Practice

PhD Thesis: Weina Jin, MD **Supervisor: Ghassan Hamarneh**

Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, Canada

Thesis Research Question

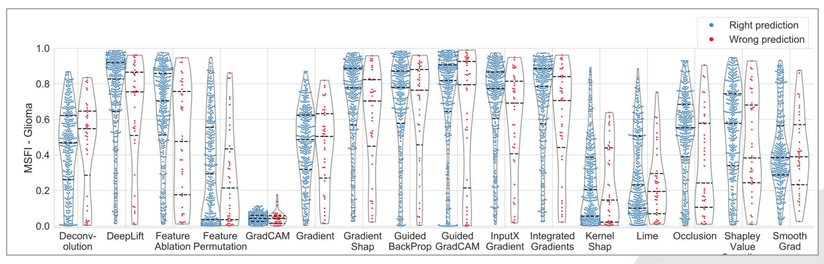
How can we—both as the collective technical community and as the individual technical developers and researchers—develop **explainable AI** (XAI) and **medical image analysis** techniques **responsibly** to genuinely benefit patients' healthcare and society?

Symptoms

Interventions

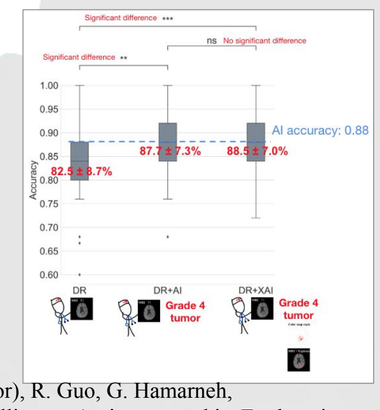
critical technical practice

Existing XAI do not fulfill clinical requirements



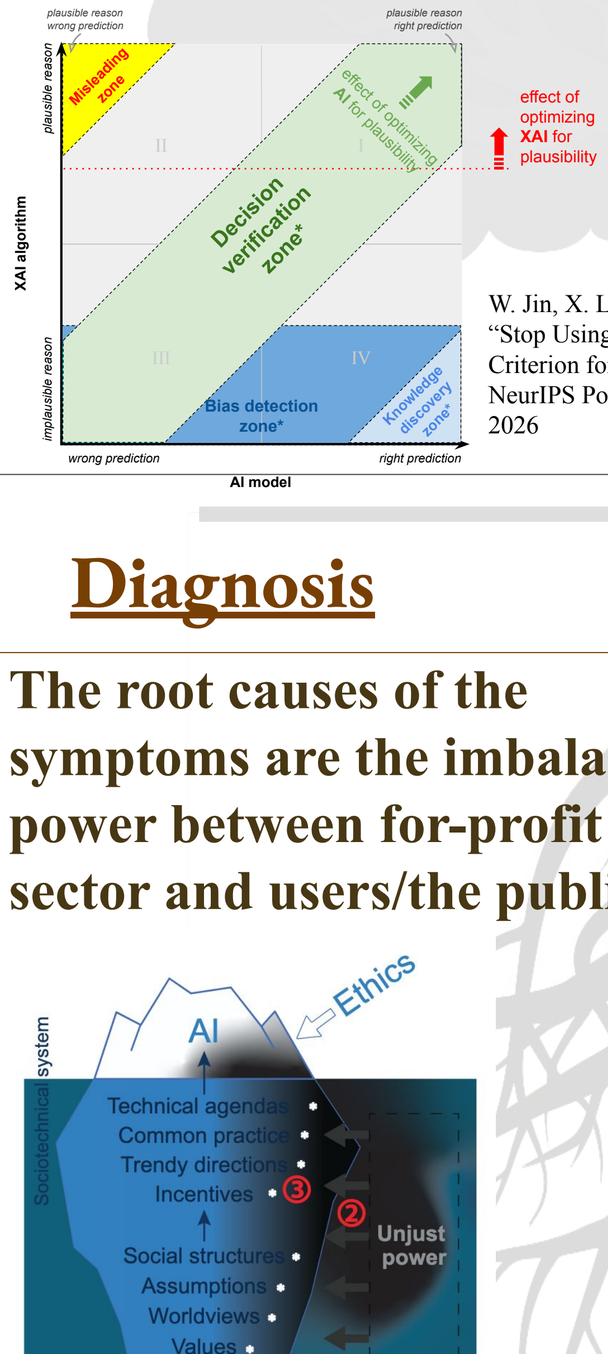
W. Jin, X. Li, G. Hamarneh, "Evaluating Explainable AI on Multi-Modal Medical Imaging Task: Can Existing Algorithms Fulfill Clinical Requirements?" AAAI 2022

Existing XAI is not helpful to improve doctors' task performance



W. Jin (co-first author), M. Fatehi (co-first author), R. Guo, G. Hamarneh, "Evaluating the Clinical Utility of Artificial Intelligence Assistance and its Explanation on the Glioma Grading Task". Artificial Intelligence in Medicine. 2024

The main XAI evaluation criterion of plausibility is unscientific and unethical



W. Jin, X. Li, G. Hamarneh, "Stop Using Plausibility as the Criterion for Explainable AI" NeurIPS Position Paper Track 2026

Diagnosis

The root causes of the symptoms are the imbalanced power between for-profit sector and users/the public

Critical analysis of AI-assisted user studies

Over-reliance can explain away the performance gain

Is the performance of User+AI better than User alone?

User + AI > User ✅

FLAWED

The AI assistance is considered to show a positive effect by helping users improve their performance

W. Jin, G. Hamarneh, "Accounting for Over-Reliance in AI-Assisted Performance Studies". MICCAI Workshop on Human-AI Collaboration. 2026

Responsible XAI framework

1. Basic properties of XAI

- ☐ **P1: Power check.** XAI is to check and shift power from the profit-oriented sector to users and society, especially the marginalized and mostly impacted communities.
- ☐ **P2: User-centred XAI.** XAI should support users' intended goals of using AI explanation and users' reasoning process.
- ☐ **P3: Bounded functionality.** The XAI/AI system has scopes, weaknesses, and limitations, some of which cannot be overcome by technical advances.
- ☐ **P4: Rejectable AI.** The associated AI system of the XAI technique and its data collection are rejectable.
- ☐ **P5: Unnecessary trade-off.** Explainability does not inevitably trade off with performance.

2. XAI design: requiring the explanation *form* to be understandable and fully explainable

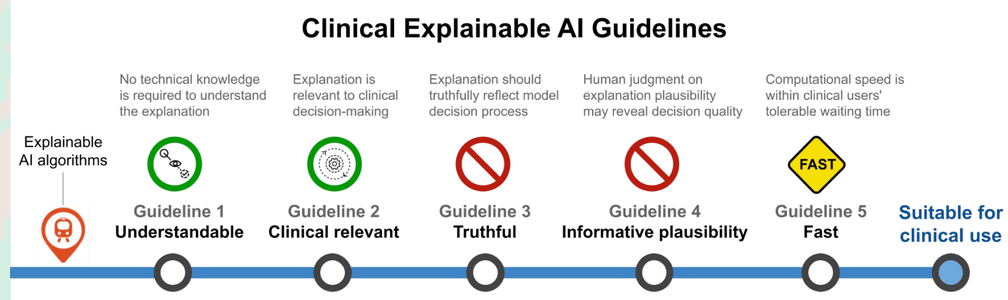
The explanation form of an XAI technique needs to be able to support each step in users' intuitive inferential reasoning of an AI explanation by providing explanatory information that can fully address the three propositions:

- ☐ ① How are the important features A identified from the input instance X?
- ☐ ② How is the important features A described and presented in human-understandable forms?
- ☐ ③ How is the output prediction Y inferred from the important features A?

3. XAI evaluation

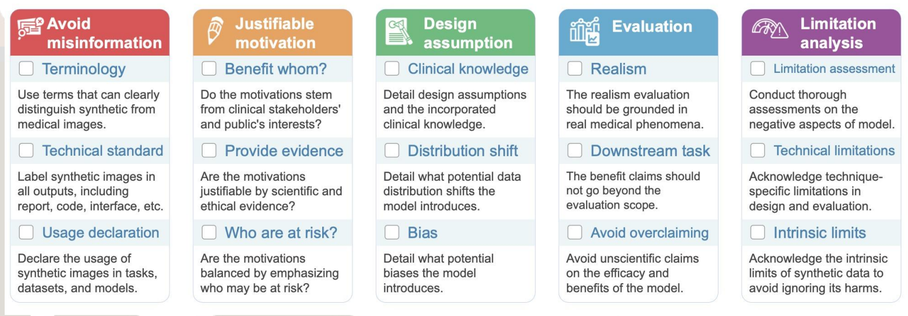
- ☐ **XAI algorithmic efficacy**: requiring the explanation *content* to be truthful and useful
  - ☐ *Truthfulness* is the primary evaluation objective for XAI.
  - ☐ *Users' intended purposes* are the secondary evaluation objective for XAI, XAI techniques can be optimized and evaluated for a specific functionality, such as verifying AI prediction, detecting biase from AI prediction, and discovering new knowledge.
- ☐ **Limitation analysis** is as important as efficacy evaluation. It includes three components:
  - ☐ **L1**. Systematically identify and assess the scope, limitations, weaknesses, failure modes, negative impacts, and risks of the XAI technique and its coupled AI system.
    - ☐ Quantify the probability of misleading explanations from the AI-XAI system on the given task.
  - ☐ **L2**. Disclose technique-specific limitations of the XAI technique and its evaluation methods.
    - ☐ [When applicable] Declare the limitation of using approximation methods (either computationally or with human subjects) to the gold standard assessment to assess the utility of XAI to users' target task.
  - ☐ **L3**. Disclose the intrinsic limits of the XAI technique and its evaluation methods.
    - ☐ [When applicable] Acknowledge the intrinsic limit of truthfulness evaluation that can only approximate, but not fully access, the underlying prediction process of a complex black-box model.

Clinical XAI guidelines

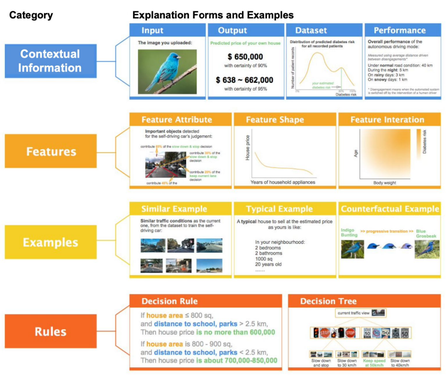


W. Jin, X. Li, M. Fatehi, G. Hamarneh, "Guidelines and Evaluation of Clinical Explainable AI in Medical Image Analysis".Medical Image Analysis 2023

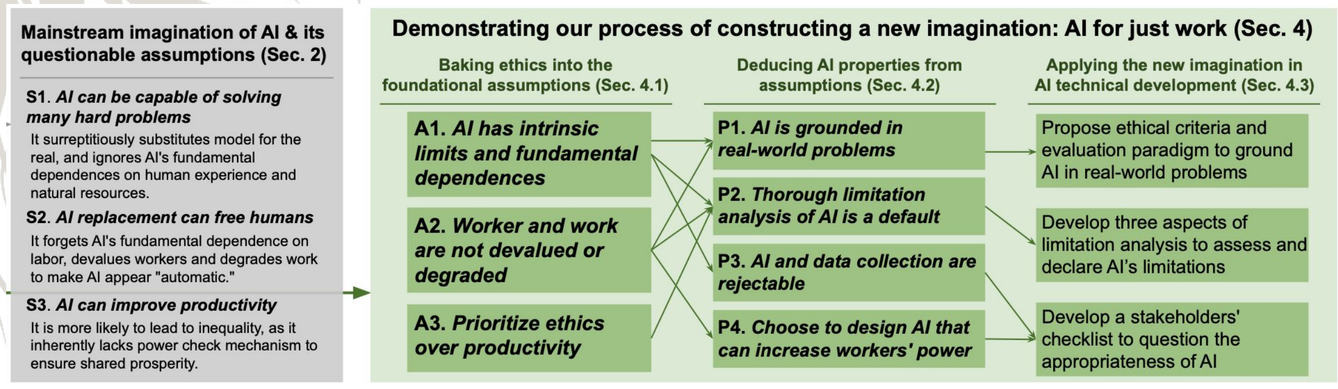
Ethical Medical Image Synthesis



End-user-centered XAI framework EUCA



Encoding justice as foundational assumptions in technical development



PhD Thesis Webpage
weina.me/phd_thesis

Video Presentation
https://youtu.be/bIUyLrWFlXc

My thesis research question:

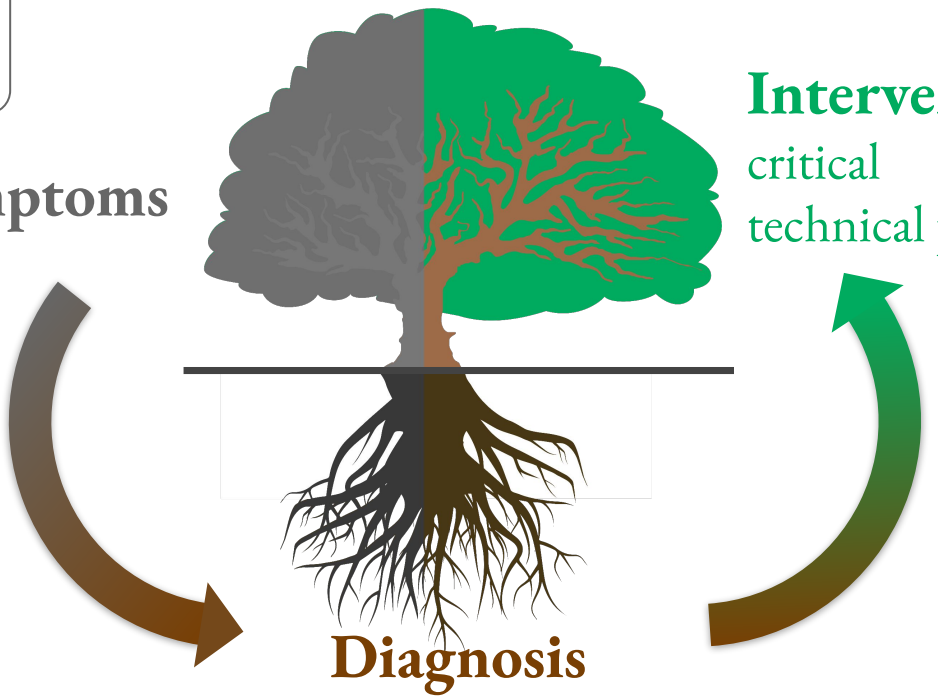
How do we develop **explainable AI (XAI)** and **medical image analysis (MIA)** techniques **responsibly** to genuinely benefit patients' healthcare and society?

Existing XAI do not fulfill clinical requirements

Existing XAI is not useful for doctors' task performance

The main XAI evaluation criterion of plausibility is unscientific and unethical

Symptoms



Diagnosis

Interventions

critical technical practice

Responsible XAI framework

Clinical XAI guidelines

End-user-centered XAI framework EUCA

Ethical Medical Image Synthesis

power asymmetry between for-profit sector and users/the public