Critical Technical Practice and Critical Communication

A case from the development of explainable AI (XAI) for medical image application

Weina Jin, MD

PhD Candidate
Medical Imaging Analysis Lab
School of Computing Science
Simon Fraser University

weina.me
weinaj@sfu.ca



SFU SIMON FRASER UNIVERSITY

HAMARNEH
Medical Image Analysis Research Group

About Weina

5th-year PhD candidate

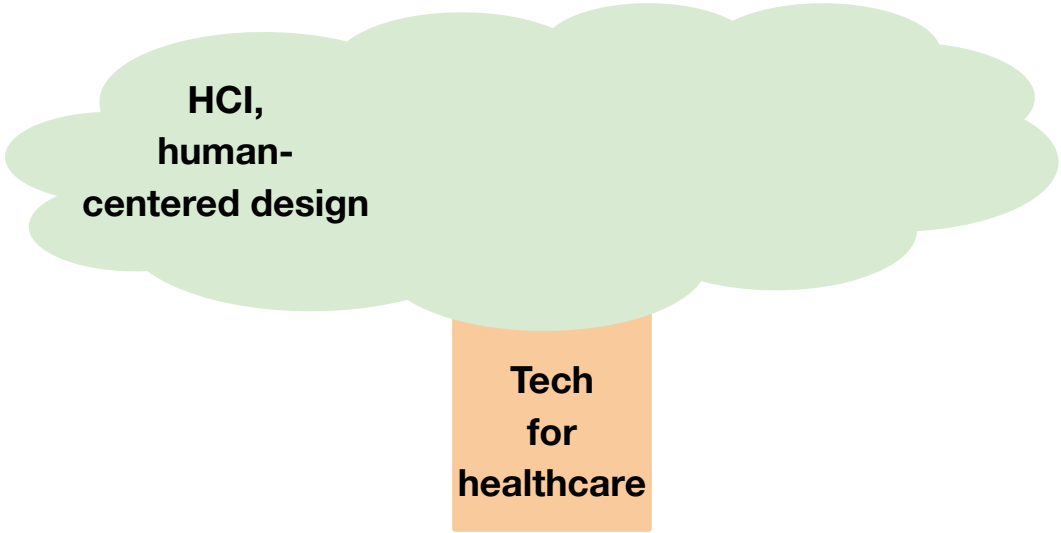
Research interest: how to make AI explainable for clinical users on medical image tasks

Tech for healthcare

About Weina

5th-year PhD candidate

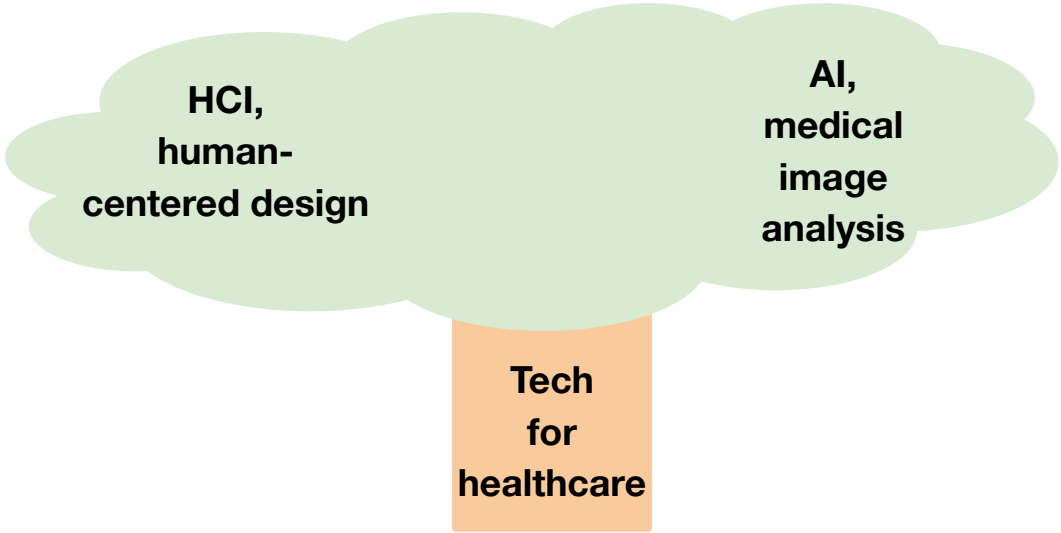
Research interest: how to make AI explainable for clinical users on medical image tasks



About Weina

5th-year PhD candidate

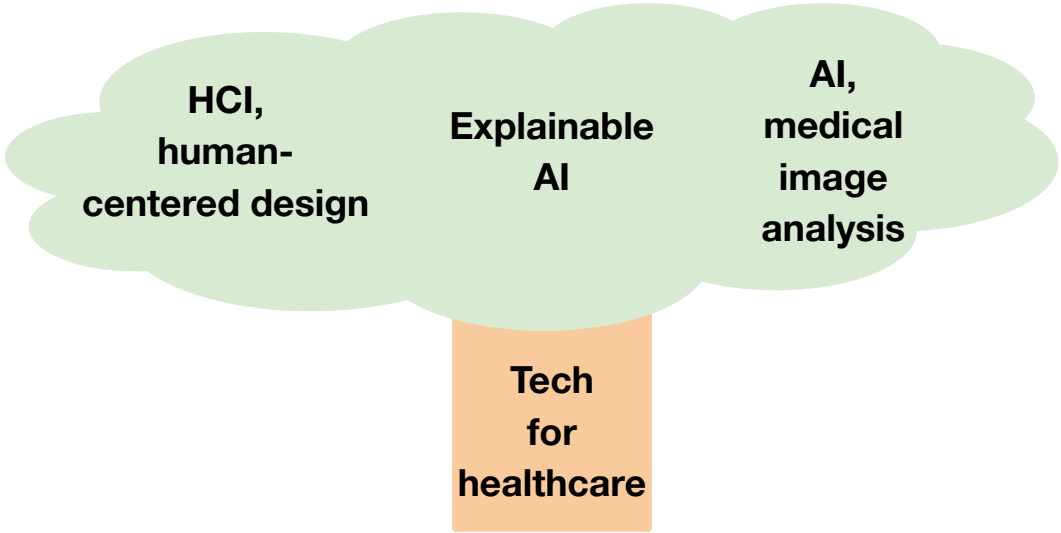
Research interest: how to make AI explainable for clinical users on medical image tasks



About Weina

5th-year PhD candidate

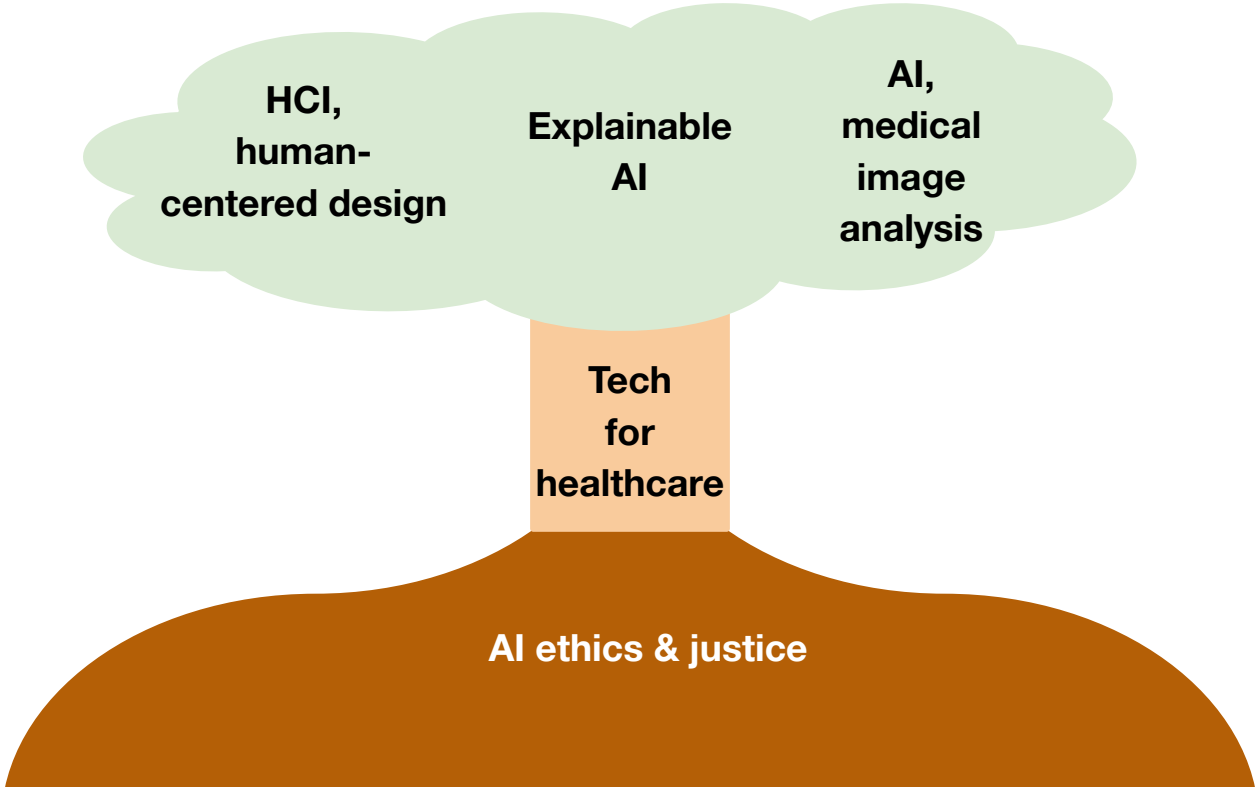
Research interest: how to make AI explainable for clinical users on medical image tasks



About Weina

5th-year PhD candidate

Research interest: how to make AI explainable for clinical users on medical image tasks

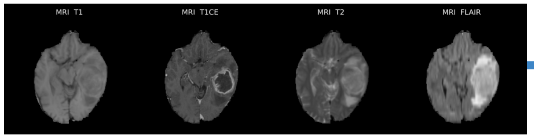


Medical image analysis: the modeling aspect

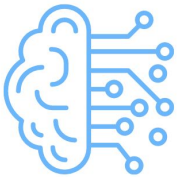
Use computational methods to interpret and analyze information from medical images

- CMPT 340 Biomedical computing
- CMPT 419 Biomedical Image Computing

For example, in clinical decision support:



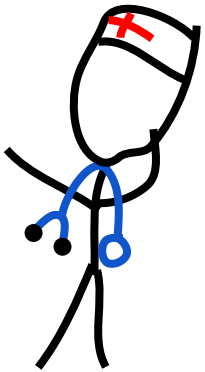
Input brain image



Model

Suggestion from the model:
High-grade brain tumor

Output

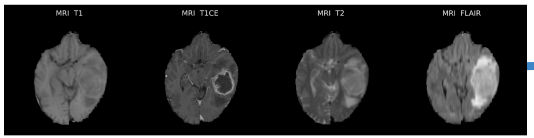


Medical image analysis: the modeling aspect

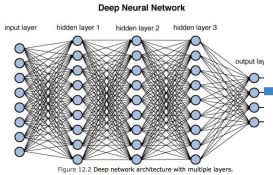
Use computational methods to interpret and analyze information from medical images

- CMPT 340 Biomedical computing
- CMPT 419 Biomedical Image Computing

For example, in clinical decision support:



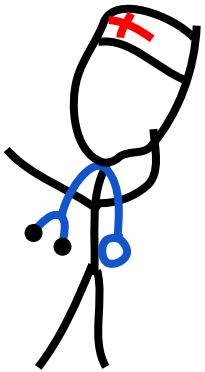
Input brain image



AI model

Suggestion from the model:
High-grade brain tumor

Output



AI models are largely black-box, human-uninterpretable models because of

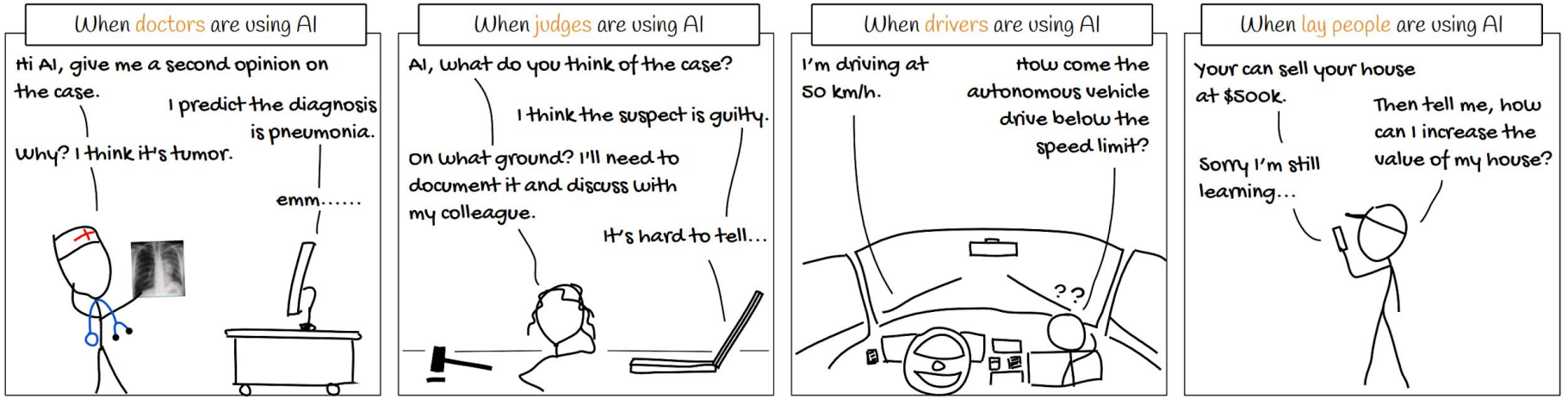
- Millions of parameters
- Many non-linear transformations from input to output

Explainable AI (XAI)

In **critical tasks** that are related to humans' life, money, career, etc.,

for humans to make informed decision assisted by a model,

the model needs to explain its decisions in human-understandable ways [1].



[1] EUCA: the End-User-Centered Explainable AI Prototyping Framework. arXiv:2102.02437
[2] Towards A Rigorous Science of Interpretable Machine Learning

Potential clinical utilities of AI explanations: the clinical users' aspect

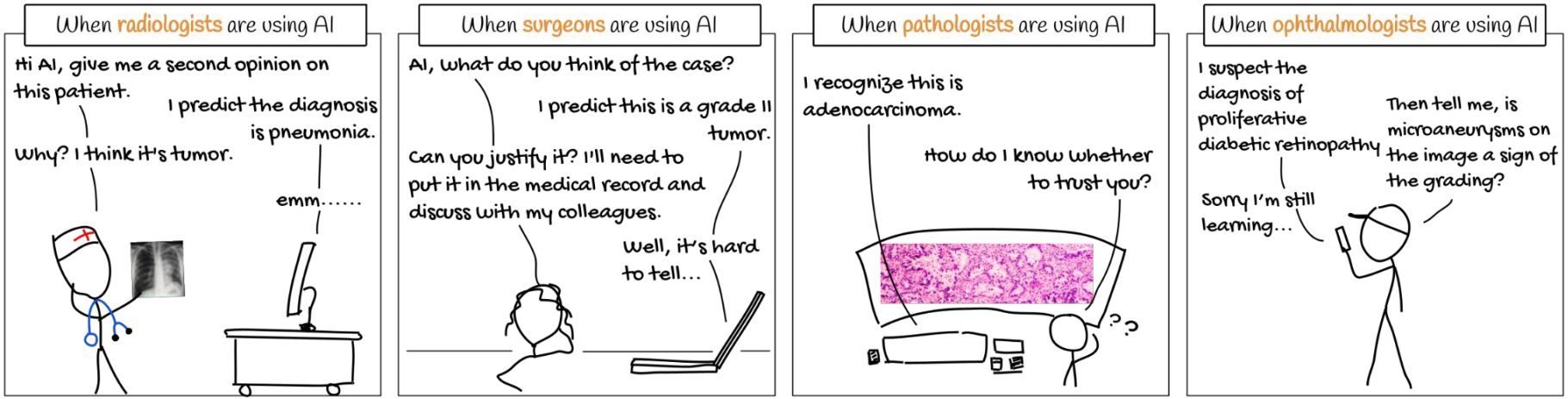
Explainable AI (XAI): To explain AI predictions in human-understandable ways [2]

Resolve clinical users' disagreement with AI [1]
informed decision

Verify AI's suggestion [1]
evidence-based medicine

Calibrate clinical users' trust [1]

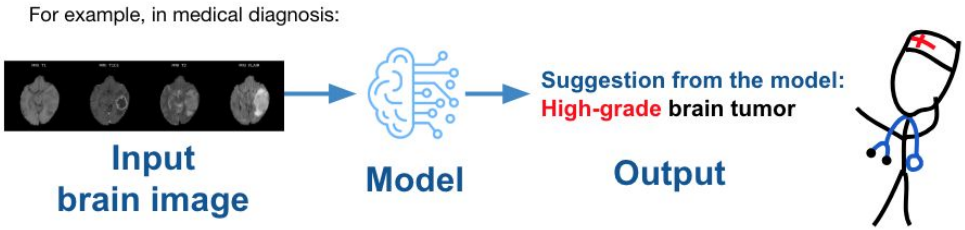
Learning & medical knowledge discovery [1]



[1] EUCA: the End-User-Centered Explainable AI Prototyping Framework. arXiv:2102.02437
[2] Towards A Rigorous Science of Interpretable Machine Learning

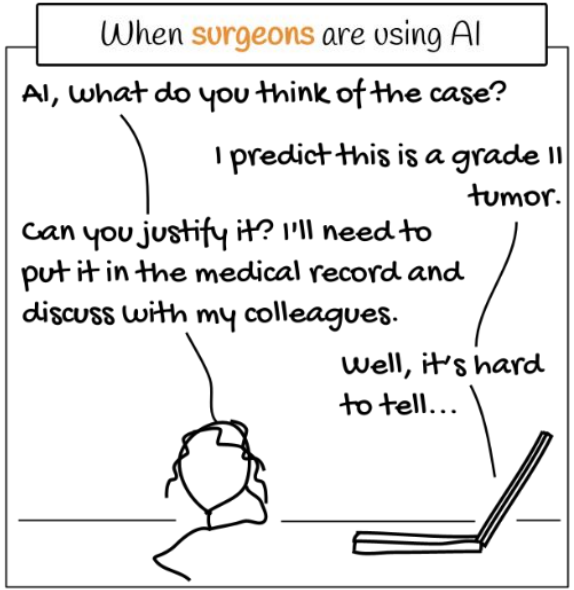
Medical image analysis

Use computational methods to interpret and analyze information from medical images



Explainable AI (XAI)

To explain AI predictions in human-understandable ways



In-class discussion and writing

A case of a clinical user study to evaluate if the XAI is helpful for doctors

Weina: Introduce the study design and result

Students: in-class discussion and writing of the result

Potential clinical utilities of AI explanations: the clinical users' aspect

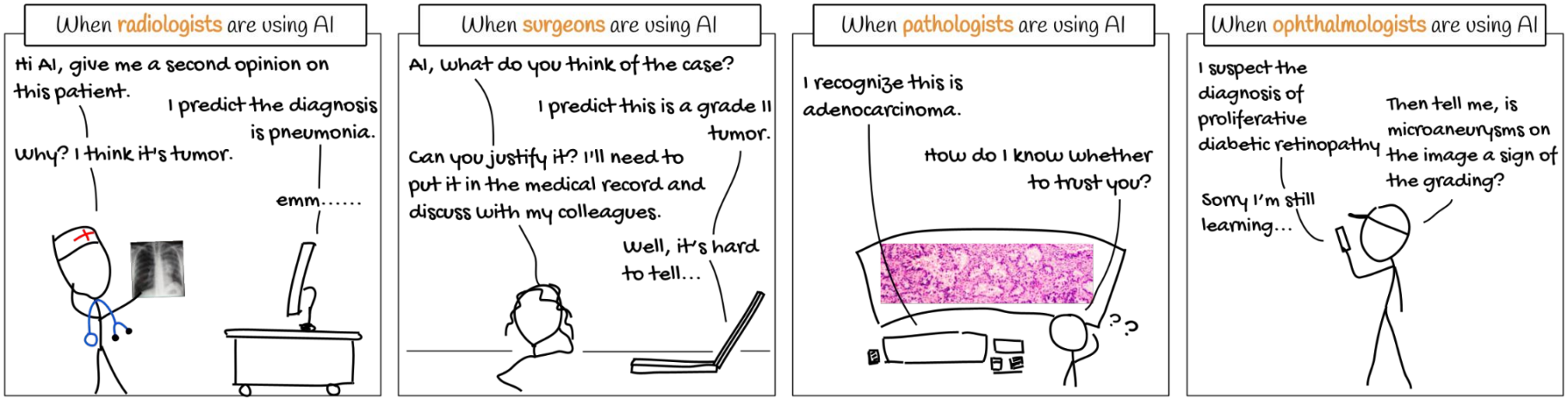
Explainable AI (XAI): To explain AI predictions in human-understandable ways [2]

Resolve clinical users' disagreement with AI [1]
informed decision

Verify AI's suggestion [1]
evidence-based medicine

Calibrate clinical users' trust [1]

Learning & medical knowledge discovery [1]



[1] EUCA: the End-User-Centered Explainable AI Prototyping Framework. arXiv:2102.02437
[2] Towards A Rigorous Science of Interpretable Machine Learning

Potential clinical utilities of AI explanations: the clinical users' aspect

Explainable AI (XAI): To explain AI predictions in human-understandable ways [2]

Resolve clinical users' disagreement with AI [1]
informed decision

Verify AI's suggestion [1]
evidence-based medicine

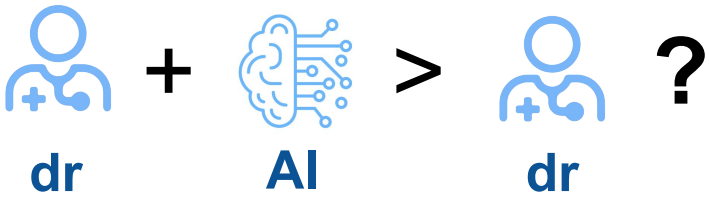
Calibrate clinical users' trust [1]

Learning & medical knowledge discovery [1]

Research question:

Is the AI assistance (prediction + explanation) helpful for clinical users?

Hypothesis:



[1] EUCA: the End-User-Centered Explainable AI Prototyping Framework. arXiv:2102.02437
[2] Towards A Rigorous Science of Interpretable Machine Learning

Study design

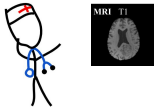
Each doctor underwent **three conditions**

Measurement: **task performance**



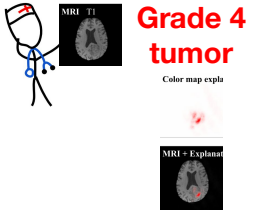
1. **DR**

Doctor only



2. **DR+AI**

Doctor assisted by AI **prediction**



3. **DR+XAI**

Doctor assisted by AI prediction & **explanation**

Research question:

Is the AI assistance (prediction + explanation) helpful for clinical users?

Hypothesis:

dr + AI > dr ?

AI: prediction, explanation

This healthcare case is relevant to similar issues when using AI to assist users

Students using AI,

HR using AI,

Lawyers using AI,

Loan officer using AI,

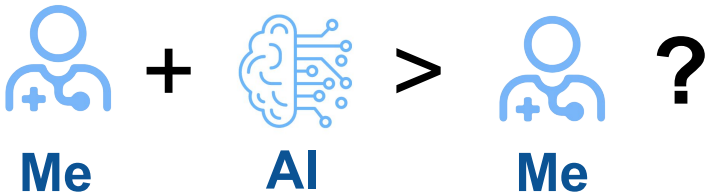
Drivers using AI,

……

Research question:

Is the AI assistance (prediction + explanation) helpful for users?

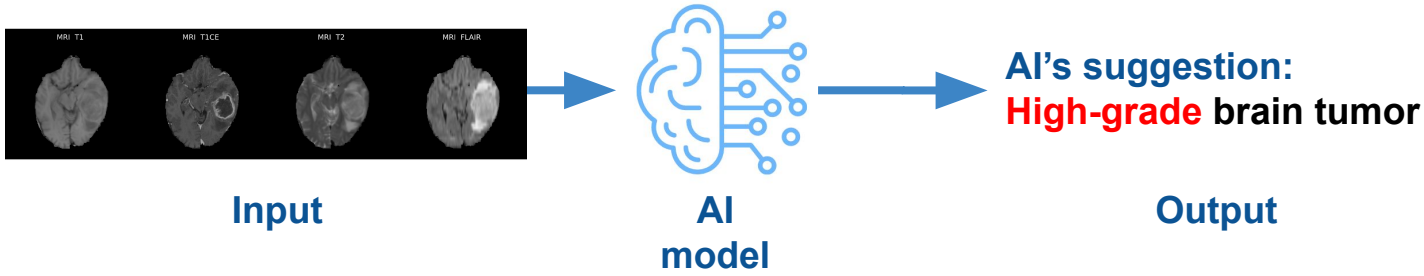
Hypothesis:



In-class writing case

Clinical task

Brain tumor grading task: recognize the brain tumor as *low* or *high* grade

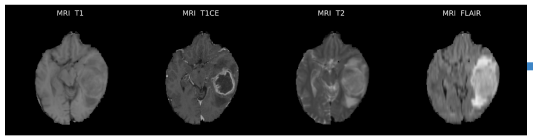


[1] Smoothgrad: removing noise by adding noise, 2017. Arxiv: 1706.03825

In-class writing case

Brain tumor grading task: recognize the brain tumor as *low* or *high* grade

AI assistance: a suggestion + an explanation



Input

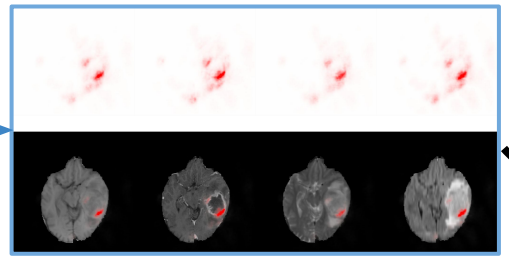
AI model

3D convolutional neural network

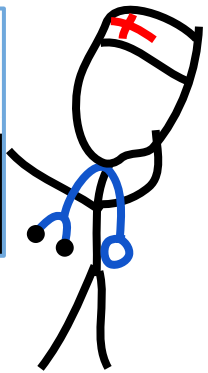
AI's suggestion:
High-grade brain tumor

Accuracy: 88%

AI's explanation:



SmoothGrad XAI algorithm [1]

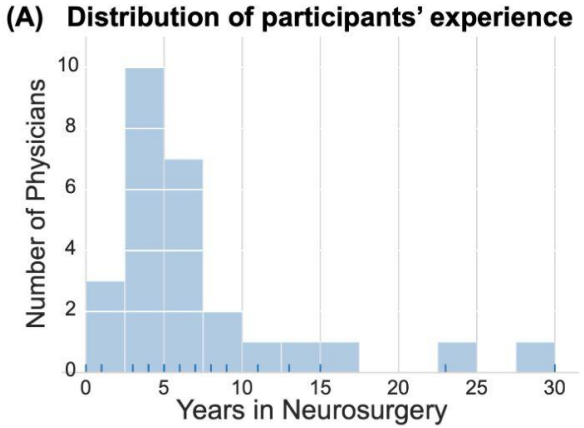


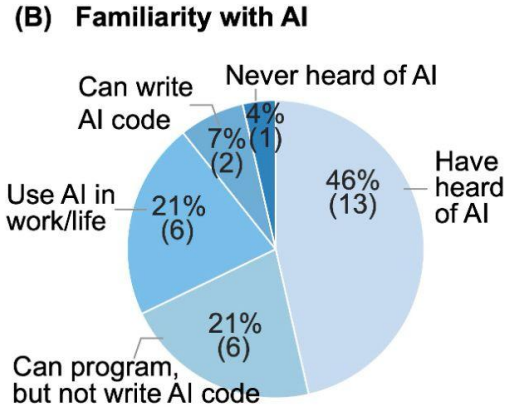
[1] Smoothgrad: removing noise by adding noise, 2017. Arxiv: 1706.03825

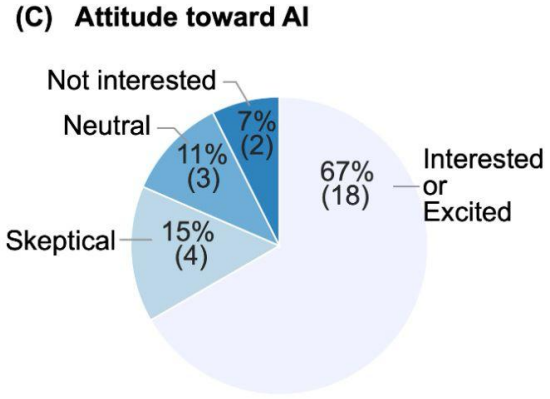
A user study on the clinical utility of AI and its explanation in a brain tumor grading task

Study Participants:

- 35 neurosurgeons
- 12 attending, 2 fellows, 21 residents
- Years of practicing neurosurgery: 7.1 ± 6.5
- Female: male = 7:19; age: 34.7±8.2







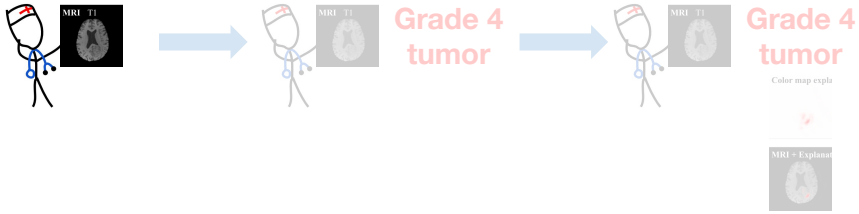
Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task. W Jin, M Fatehi, R Guo, G Hamarneh. AI in Medicine. 2024

In-class writing case

Study procedure

35 neurosurgeons, each read 25 brain images

Drs gave judgment on three conditions:



1. **DR**

Doctor only

2. **DR+AI**

Doctor assisted by AI **prediction**

3. **DR+XAI**

Doctor assisted by AI prediction & **explanation**



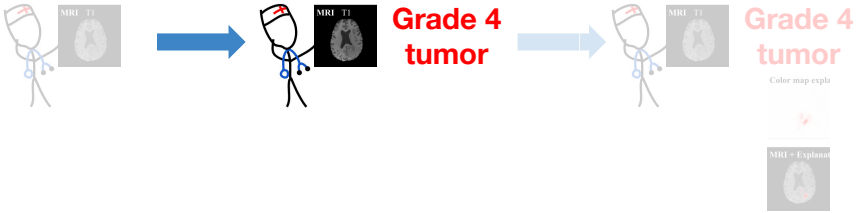
Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task. W Jin, M Fatehi, R Guo, G Hamarneh. AI in Medicine. 2024

In-class writing case

Study procedure

35 neurosurgeons, each read 25 brain images

Drs gave judgment on three conditions:



1. DR

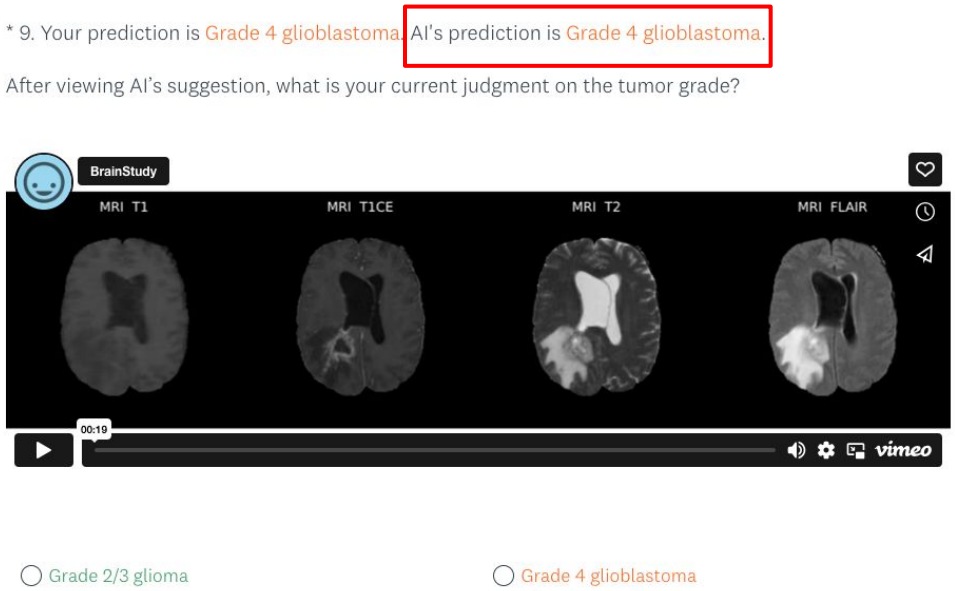
Doctor only

2. **DR+AI**

Doctor assisted by AI **prediction**

3. DR+XAI

Doctor assisted by AI prediction & **explanation**



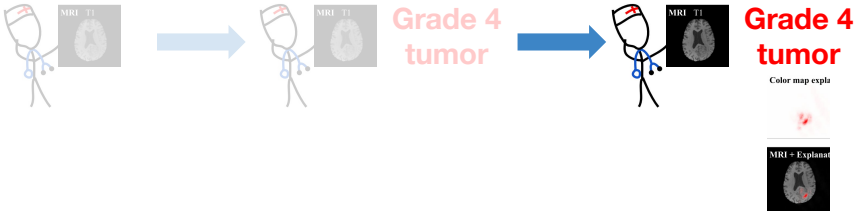
Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task. W Jin, M Fatehi, R Guo, G Hamarneh. AI in Medicine. 2024

In-class writing case

Study procedure

35 neurosurgeons, each read 25 brain images

Drs gave judgment on three conditions:



1. **DR**

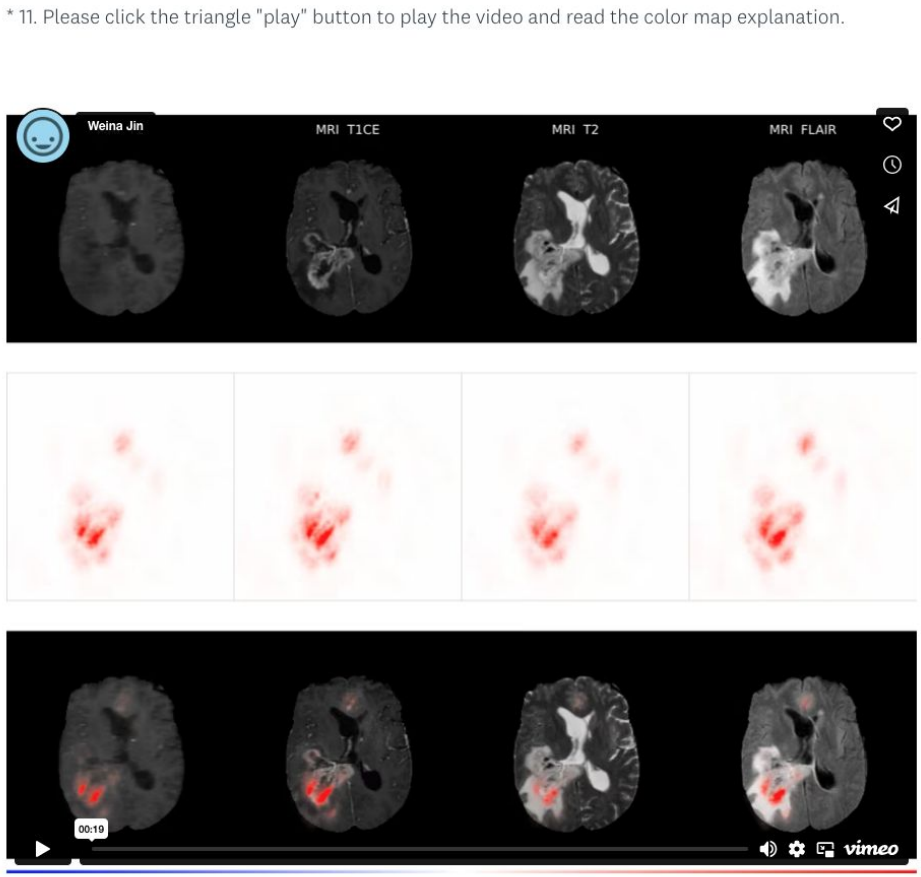
Doctor only

2. **DR+AI**

Doctor assisted by AI **prediction**

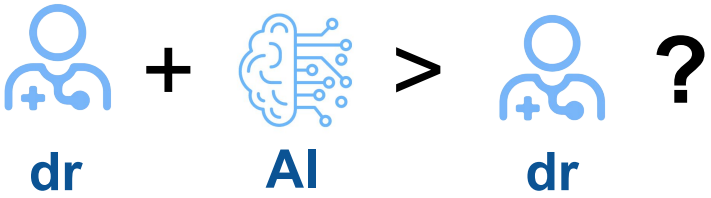
3. **DR+XAI**

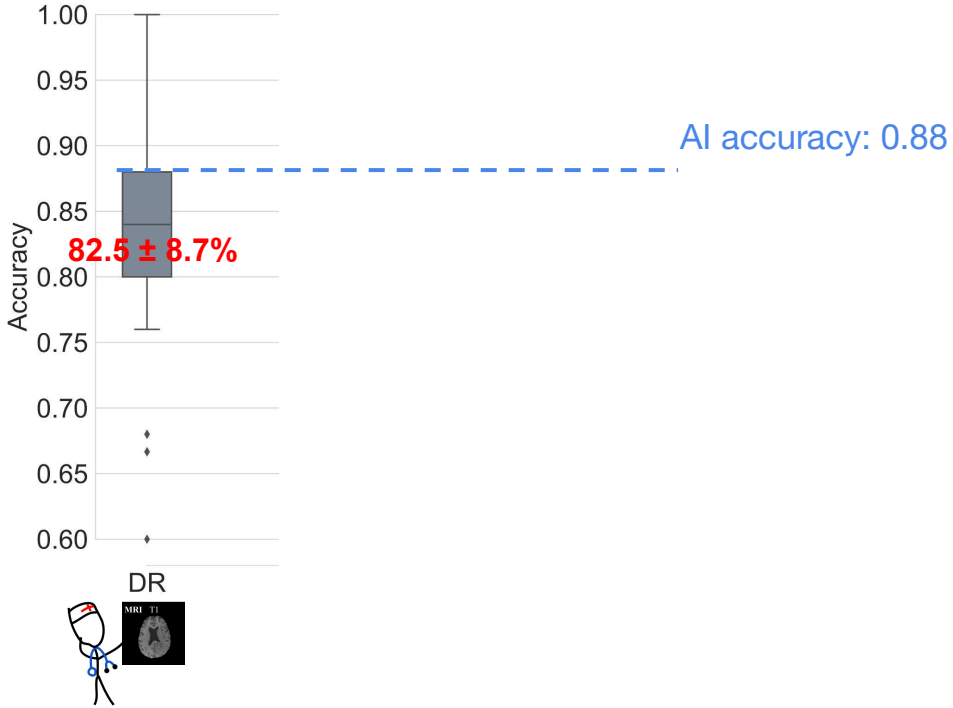
Doctor assisted by AI prediction & **explanation**



Result

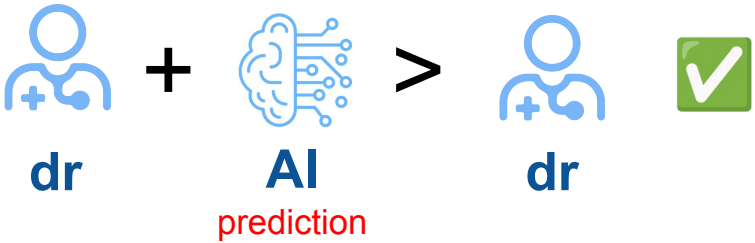
Is doctor+AI better than doctor alone?

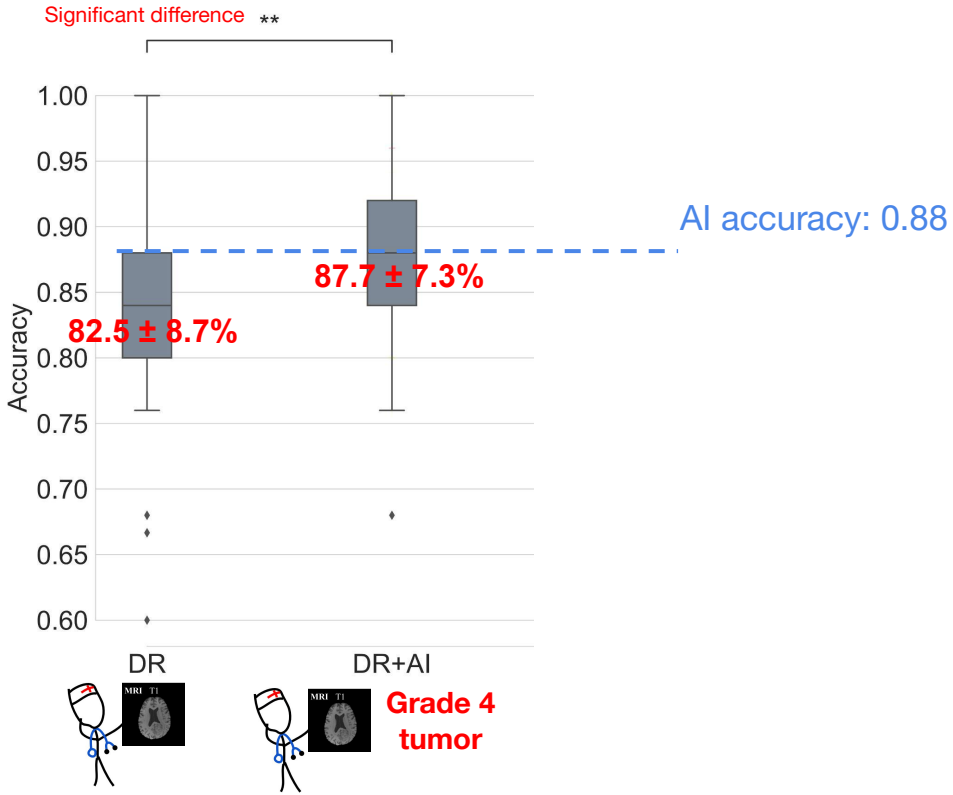




Result

Is doctor+AI better than doctor alone?





In an experimental setting:

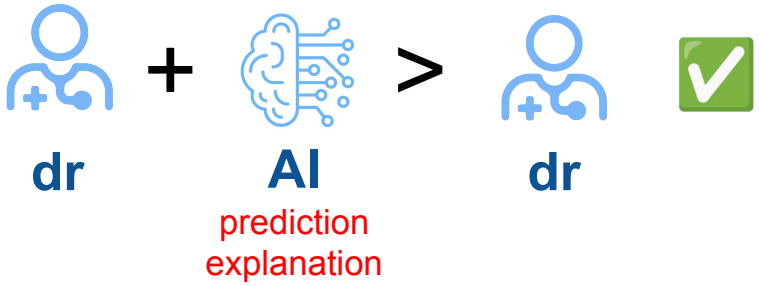
- AI prediction (**DR+AI**) is helpful to improve doctors' task performance

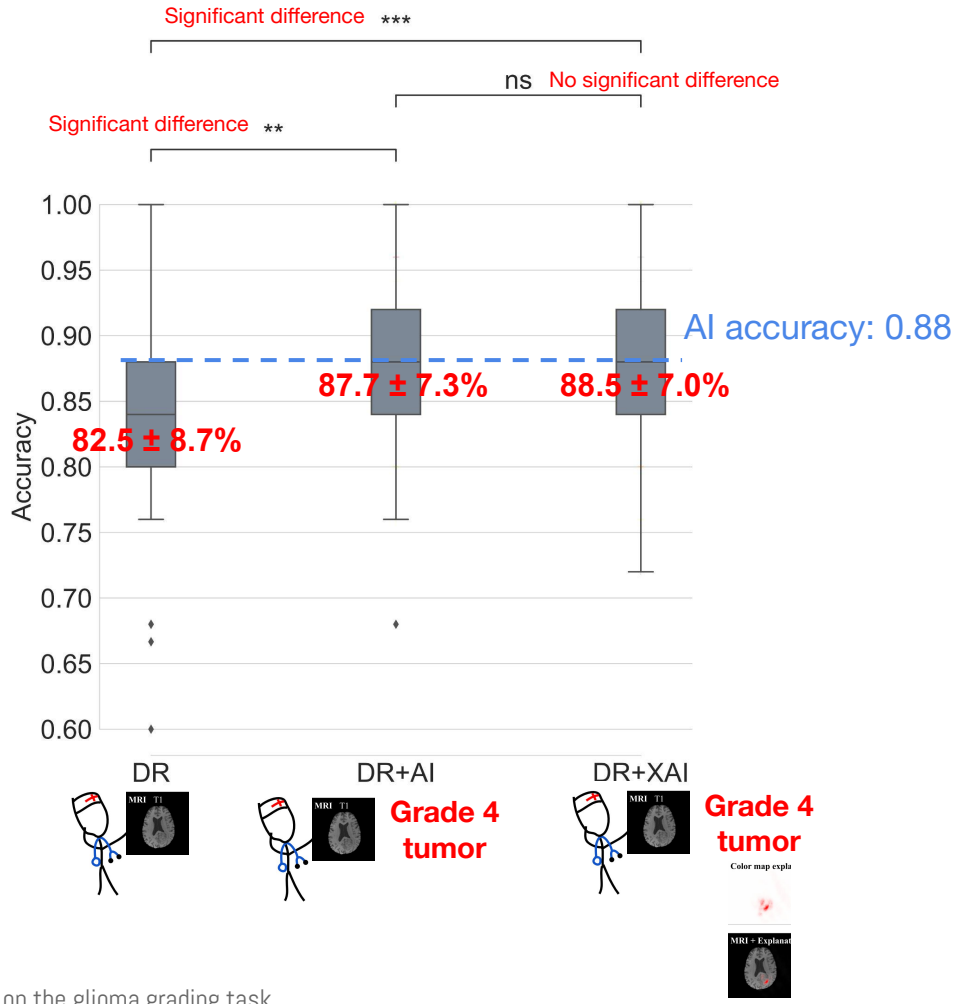
Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task. W Jin, M Fatehi, R Guo, G Hamarneh. AI in Medicine. 2024

In-class writing case

Result

Is doctor+AI better than doctor alone?





In an experimental setting:

- AI prediction (**DR+AI**) is helpful to improve doctors' task performance
- AI explanation (**DR+XAI**) does not show additional help in improving doctors' performance

Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task.

How would you interpret and communicate the study result to your audience?

Four roles

Role 1: Doctor

A medical intern writing to a junior doctor advisory committee

Role 2: Patient

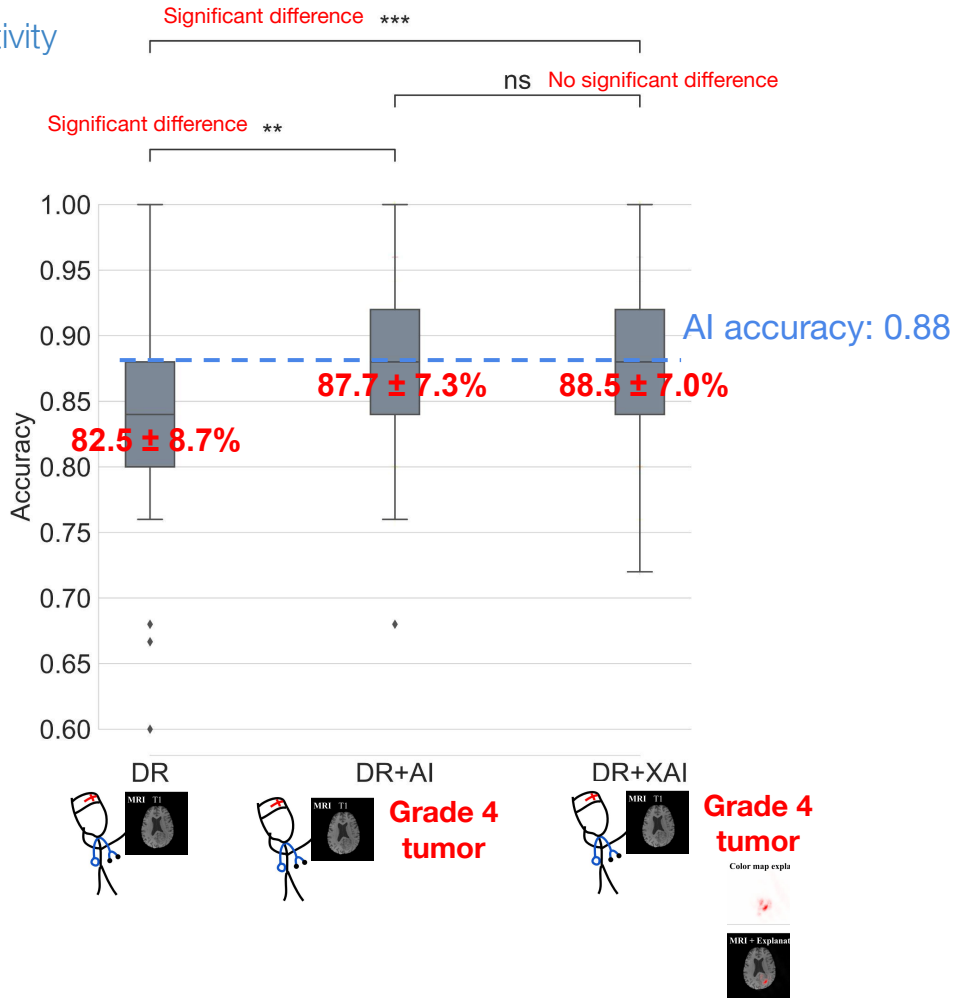
A patient suspecting malpractice writing to advocacy group

Role 3: Hospital

A hospital IT administrator writing to the hospital staff and management team

Role 4: Technology company

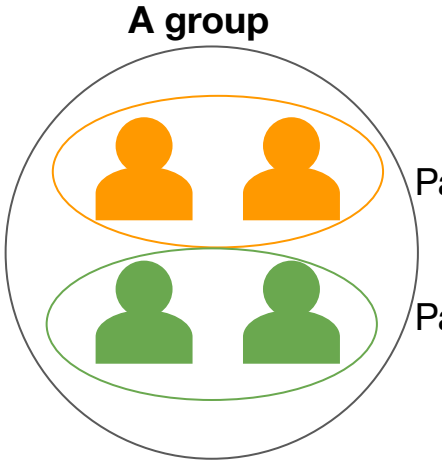
A manager of a development team of the AI system writing to the company's upper management team



ACTIVITY Pair discussion + Individual writing 1 + Group discussion + Individual writing 2

Four roles

Role 1: Doctor A medical intern writing to a junior doctor advisory committee	**Role 2: Patient** A patient suspecting malpractice writing to advocacy group
Role 3: Hospital A hospital IT administrator writing to the hospital staff and management team	**Role 4: Technology company** A manager of a development team of the AI system writing to the company's upper management team

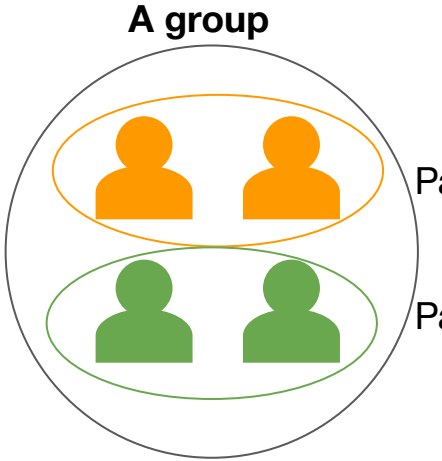


1. Form a **4-person group** with neighboring students
2. **Each group** is divided into **two pairs** of students.
3. Each **pair selects a role** that is different from the other pair.

ACTIVITY Pair discussion + Individual writing 1 + Group discussion + Individual writing 2

Four roles

Role 1: Doctor A medical intern writing to a junior doctor advisory committee	**Role 2: Patient** A patient suspecting malpractice writing to advocacy group
Role 3: Hospital A hospital IT administrator writing to the hospital staff and management team	**Role 4: Technology company** A manager of a development team of the AI system writing to the company's upper management team



4. [5 min Pair discussion] brainstorm and exchange ideas, list notes and key points of your recommendation according the chosen role (next slide)

5. [5 min Individual writing 1] Then, write your paragraphs **individually**

6. [10 min Group discussion + writing 2]

Pair Discussion How would you interpret and communicate the study result?

1. Based on your role (details in the next slide), discuss the interpretation of the study result

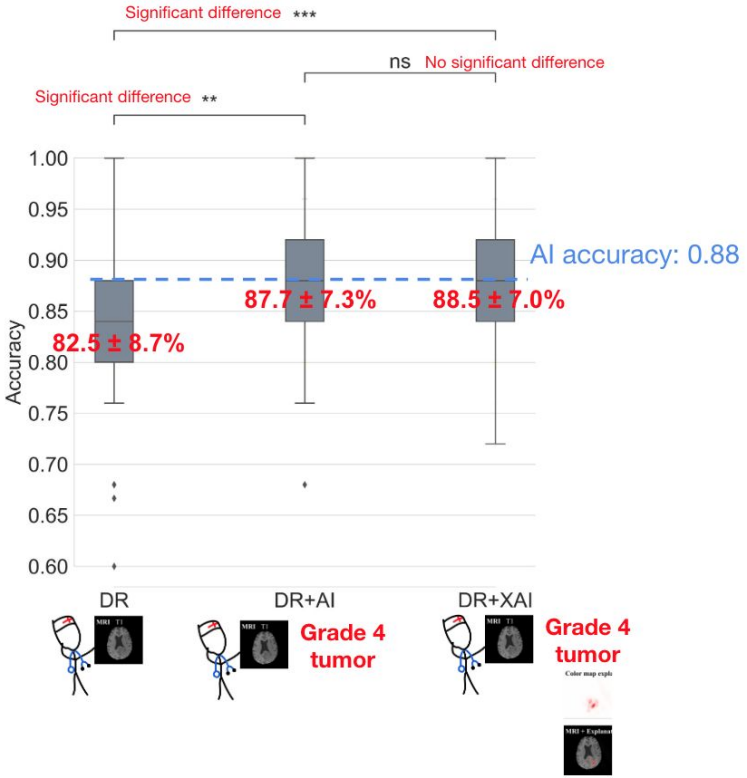
2. In your discussion, you need to specify that you recommend (a), (b), or (c) in clinical settings based on your interpretation of the study evidence

   a). **No AI**

   b). **AI prediction** only, i.e., a black-box AI that only provides predictions/conclusions to users

   c). **AI prediction + explanation**

3. Brainstorm and exchange ideas, list notes and key points. The notes will be the starting point for your individual writing next.



Individual Writing 1 How would you interpret and communicate the study result?

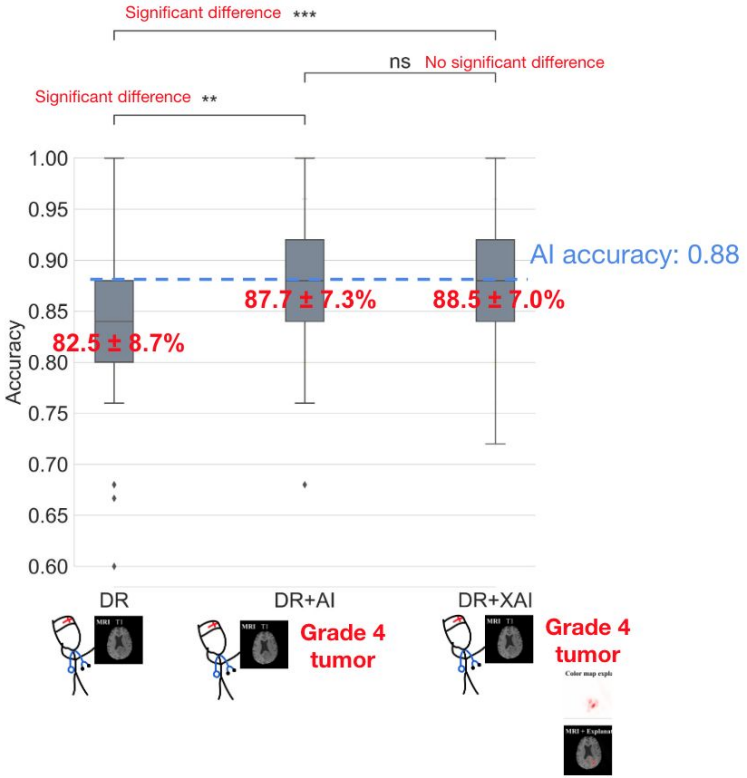
1. Based on your role (details in the next slide), write a few paragraphs to communicate your interpretation of the study result to your target audience

2. In your writing, you need to specify that you recommend (a), (b), or (c) in clinical settings based on your interpretation of the study evidence

   a). **No AI**

   b). **AI prediction** only, i.e., a black-box AI that only provides predictions/conclusions to users

   c). **AI prediction + explanation**



Roles in more detail

How would you interpret and communicate the study result?

Role 1: Doctor

Suppose you are a **medical intern** serving on a junior doctor advisory committee to a hospital management team. Recently, the hospital is looking into purchasing a potential AI clinical decision support system, and would like to seek advice from the junior doctor advisory committee to understand the influence of such an AI system on junior doctors.

Please **write to the junior doctor advisory committee** to share your opinions.

Role 2: Patient

Suppose you are a **patient** who is experiencing medical malpractice. And you suspect that part of the reason for the malpractice is due to the doctor overrelying on the AI-based clinical decision system. You have been shown the evidence from this study of the efficacy of the AI system the hospital uses.

Please **write to a patient advocacy group** to communicate your interpretation of the study result.

Role 3: Hospital

Suppose you are the **IT administrator/data scientist of a hospital**. You are collecting and reviewing information to inform the stakeholders about the decision of whether to purchase a potential AI clinical decision support system.

Please **write to stakeholders** (hospital staff and management team) to facilitate their decision.

Role 4: Technology company

Suppose you are the **manager of a development team** in a large tech company, working on an AI clinical decision support system (think of IBM Watson or Google Health). A large hospital is interested in your product and has just started the inquiry process.

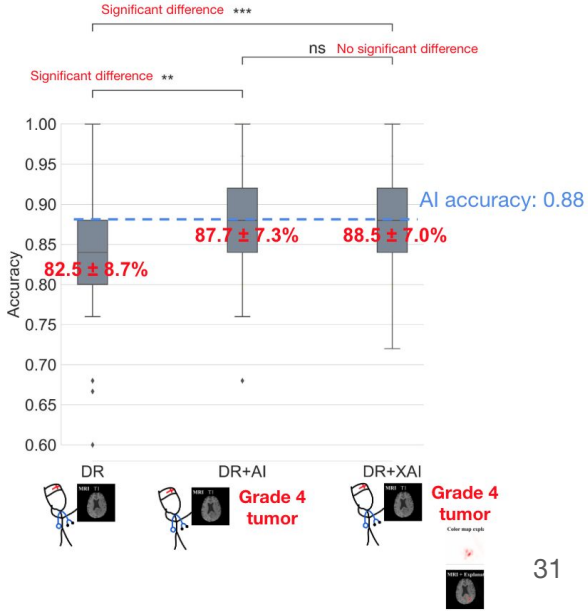
Please **write to your upper management team**, stating your position and arguments regarding your interpretation of the result and how result from this study can influence your product.

Based on the role, interpret the study result, and use the study evidence to provide rationales for your recommendations for (a), (b), or (c) in clinical settings:

a). **No AI**

b). **AI prediction** only, i.e., a black-box AI that only provides predictions/conclusions to users

c). **AI explanation** for its prediction



Job descriptions in more detail

Role 3: Hospital
Suppose you are the **IT administrator/data scientist of a hospital**. You are collecting and reviewing information to inform the stakeholders about the decision of purchasing a potential AI clinical decision support system.

IT administrator/data scientist of a hospital

1. Participate in process reviews, workflow analysis and redesign.
2. Develop Request for Proposal documents, collaborating with users to define functional specifications and **leading the project team through the evaluation/selection of computer applications**.
3. Provide expertise in the evaluation and/or programming of computer applications, leading assigned teams to facilitate implementation of applications.
4. Coordinate the transition to live environment of business/clinical applications.
5. Install, configure, test, maintain, monitor, and troubleshoot business application systems.
6. Serve as a resources person, providing consultation and timely support to user groups.

From https://sah.on.ca/careers/information-systems-administrator-information-technology-full-time/

Role 4: Technology company
Suppose you are the **manager of a development team** in a large tech company, working on an AI clinical decision support system (think of IBM Watson or Google Health). A large hospital is interested in your product and has just started the inquiry process.

Software Engineer Manager

1. **Set and communicate team priorities** that support the broader organization's goals. Align strategy, processes, and decision-making across teams.
2. Develop the mid-term technical goal and roadmap within the scope of your teams. Evolve the roadmap to meet anticipated future requirements and infrastructure needs.
3. **Design, guide and vet systems designs** within the scope of the broader area, and write product or system development code to solve ambiguous problems.
4. Review code developed by other engineers and provide feedback to ensure best practices (e.g., style guidelines, checking code in, accuracy, testability, and efficiency).

From https://www.google.com/about/careers/applications/jobs/results/131097801969803974-software-engineer-manager-ii/

Group Discussion

How does the role assignment affect your interpretation and communication of the study result?

Discuss and compare your recommendation, interpretation, and rationale with your whole group members of 4 students who are in two different roles

1. Do you have the same or different recommendations from this study?
2. Does the sameness or difference depend on the role you have?
3. After exchanging arguments, do your opinion and recommendation change?
4. Do you notice any differences in the two discussion dynamics, i.e., people from the same vs. different roles/perspectives?
5. What do you learn from this activity?

Role 1: Doctor

A medical intern writing to a junior doctor advisory committee

Role 2: Patient

A patient suspecting malpractice writing to advocacy group

Role 3: Hospital

A hospital IT administrator writing to the hospital staff and management team

Role 4: Technology company

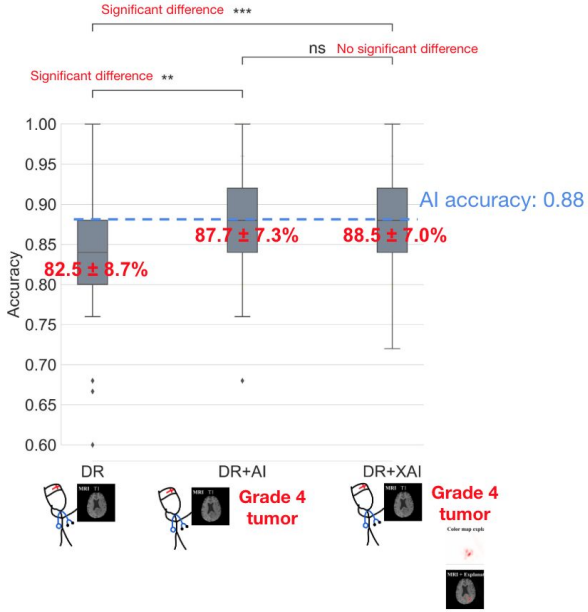
A manager of a development team of the AI system writing to the company's upper management team

Based on the role, interpret the study result, and use the study evidence to provide rationales for your recommendations for (a), (b), or (c) in clinical settings:

a). **No AI**

b). **AI prediction** only, i.e., a black-box AI that only provides predictions/conclusions to users

c). **AI explanation** for its prediction



Individual Writing 2

How does the role assignment affect your interpretation and communication of the study result?

Write a few paragraphs to reflect your thinking of one or several of the following questions:

1. Do you and your group members have the same or different recommendations from this study?
2. Does the sameness or difference depend on the role you have?
3. After exchanging arguments, do your opinion and recommendation change?
4. Do you notice any differences in the two discussion dynamics, i.e., people from the same vs. different roles/perspectives?
5. What do you learn from this activity?

Role 1: Doctor

A medical intern writing to a junior doctor advisory committee

Role 2: Patient

A patient suspecting malpractice writing to advocacy group

Role 3: Hospital

A hospital IT administrator writing to the hospital staff and management team

Role 4: Technology company

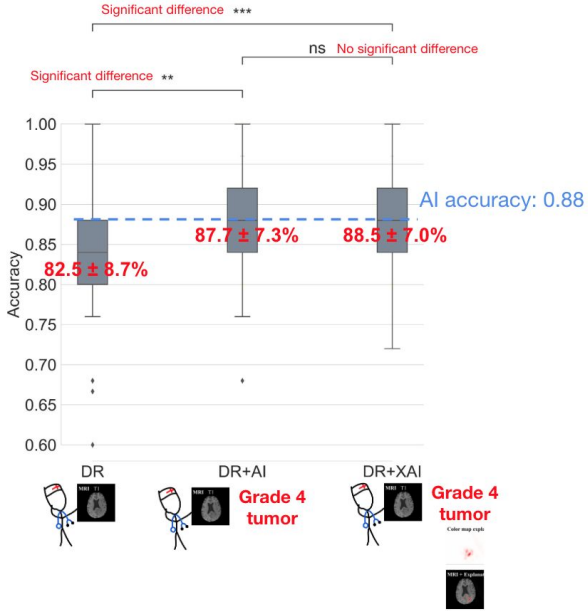
A manager of a development team of the AI system writing to the company's upper management team

Based on the role, interpret the study result, and use the study evidence to provide rationales for your recommendations for (a), (b), or (c) in clinical settings:

a). **No AI**

b). **AI prediction** only, i.e., a black-box AI that only provides predictions/conclusions to users

c). **AI explanation** for its prediction

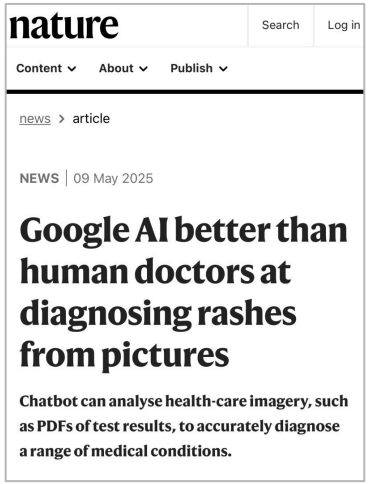


What ideas emerged during the discussions?

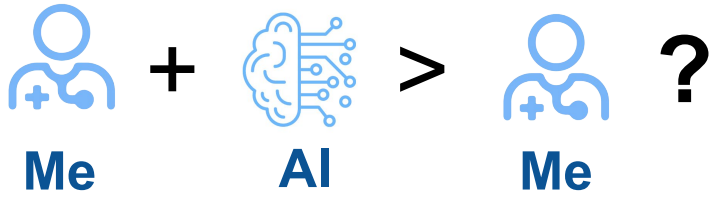
This activity is relevant to similar issues when using AI to assist users

Similar findings and claims on

AI outperforms humans in x, y, z, … tasks







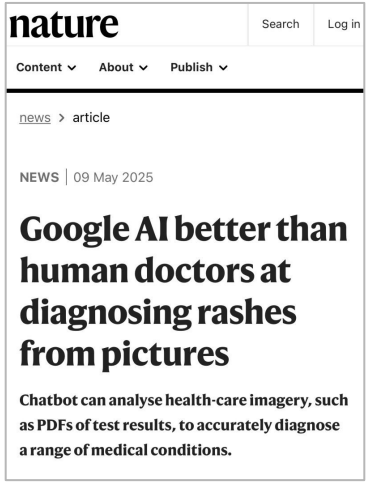
prediction,explanation

This exercise is relevant to similar issues when using AI to assist users

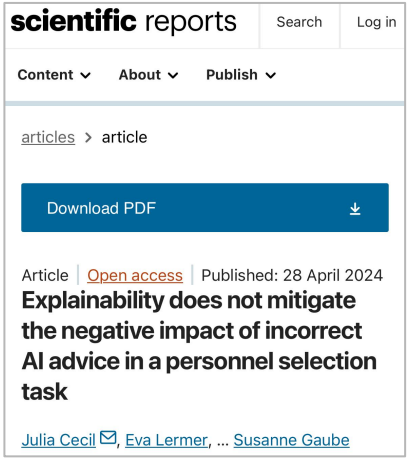
Similar findings and claims on

AI outperforms humans in a, b, c, … tasks

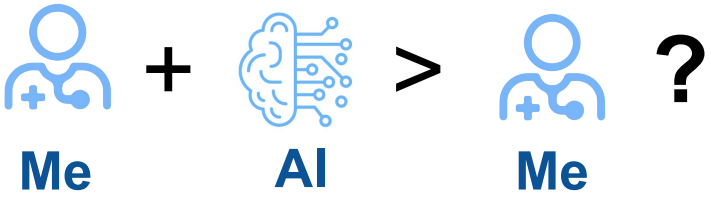
AI explanation fails in x, y, z, … tasks







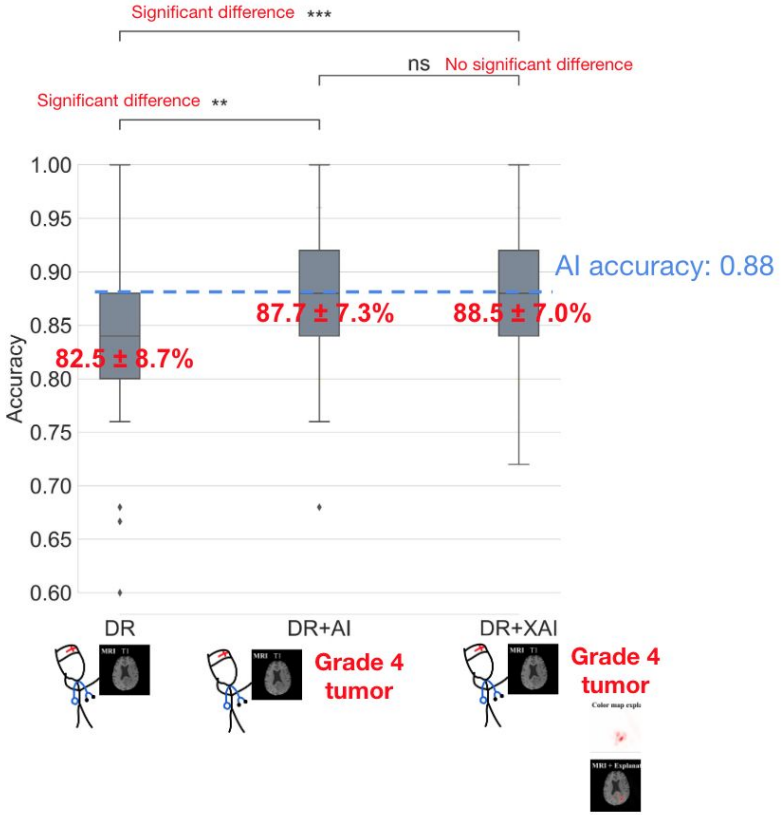




Our story of identifying different interpretations of the result

AI optimism

1. Black-box AI is **superior** to doctors;
2. Explainable AI is **useless**.
3. So we can **replace doctors with black-box AI** to improve medical decisions.



Making Power Explicable in AI: Analyzing, Understanding, and Redirecting Power to Operationalize Ethics in AI Technical Practice.
Weina Jin, Elise Li Zheng, Ghassan Hamarneh. 2025. arXiv:2510.10588

Critical thinking & communication

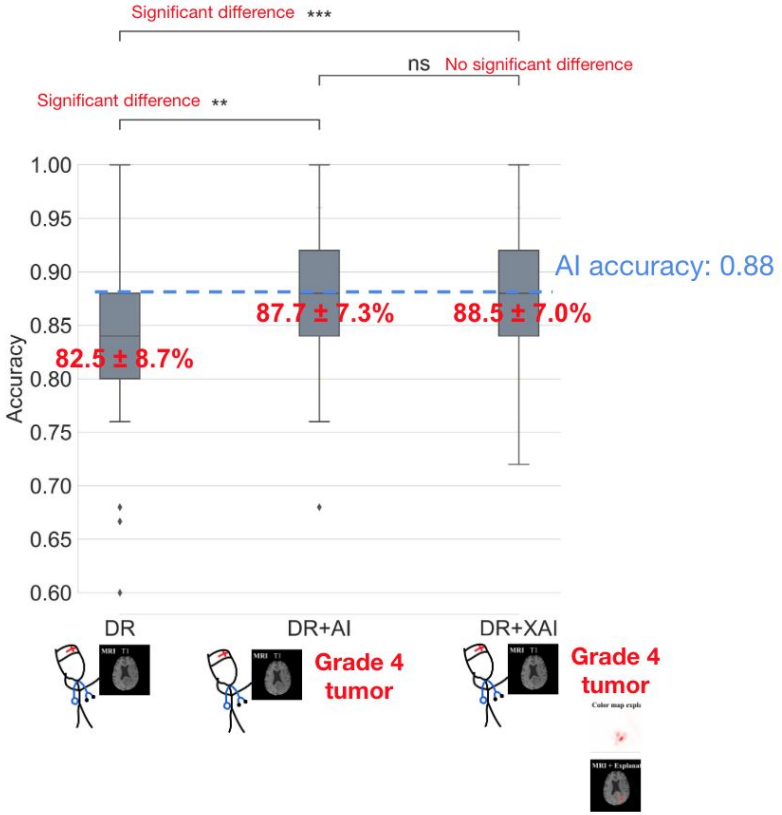
The **same evidence** can be interpreted differently to support **opposite recommendations**

AI optimism

1. Black-box AI is **superior** to doctors;
2. Explainable AI is **useless**.
3. So we can **replace doctors with black-box AI** to improve medical decisions.

AI skepticism

1. This is a **small sample** study in an experimental setting. The result needs to be **further verified** in large-scale, real-world tasks.
2. Other factors can influence/explain away the results, such as **overreliance** on AI, and **weakness in XAI**.



Making Power Explicable in AI: Analyzing, Understanding, and Redirecting Power to Operationalize Ethics in AI Technical Practice.
Weina Jin, Elise Li Zheng, Ghassan Hamarneh. 2025. arXiv:2510.10588

Critical thinking & communication

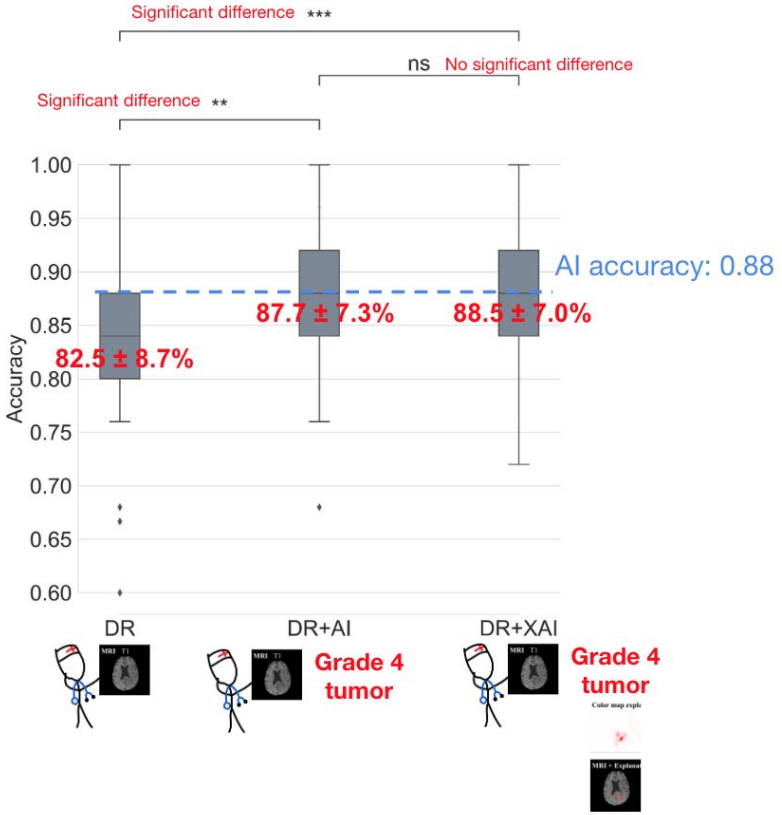
The **same evidence** can be interpreted differently to support **opposite recommendations**

AI optimism → Develop AI to **compete** with humans

1. Black-box AI is **superior** to doctors;
2. Explainable AI is **useless**.
3. So we can **replace doctors with black-box AI** to improve medical decisions.

AI skepticism → Develop AI to **collaborate** with humans

1. This is a **small sample** study in an experimental setting. The result needs to be **further verified** in large-scale, real-world tasks.
2. Other factors can influence/explain away the results, such as **overreliance** on AI, and **weakness in XAI**.



Making Power Explicable in AI: Analyzing, Understanding, and Redirecting Power to Operationalize Ethics in AI Technical Practice.
Weina Jin, Elise Li Zheng, Ghassan Hamarneh. 2025. arXiv:2510.10588

Critical thinking & communication

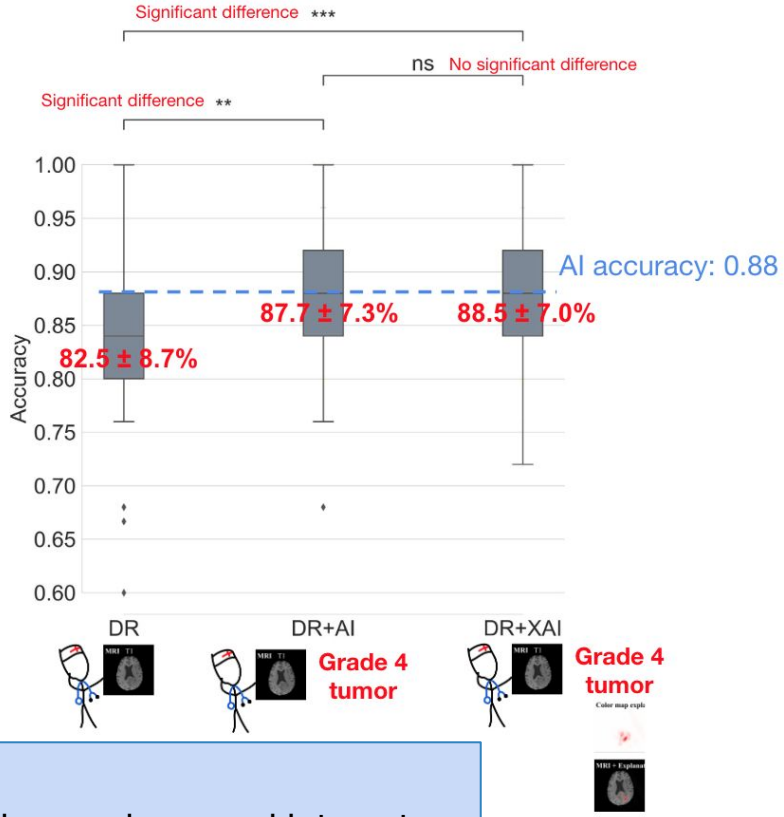
The **same evidence** can be interpreted differently to support **opposite recommendations**

AI optimism → Develop AI to **compete** with humans

1. Black-box AI is **superior** to doctors;
2. Explainable AI is **useless**.
3. So we can **replace doctors with black-box AI** to improve medical decisions.

AI skepticism → Develop AI to **collaborate** with humans

1. This is a **small sample** study in an experimental setting. The result needs to be **further verified** in large-scale, real-world tasks.
2. Other factors can influence/explain away the results, such as **overreliance** on AI, and **weakness**



Take home message:
A piece of communication implies some goals, perspectives, values, and interests behind it, though some may appear to be scientific, objective, or value-neutral [1].

[1] Feminist Epistemology and Philosophy of Science: An Introduction. Book by Kristen Intemann and Sharon Crasnow

Critical thinking & communication

The **same evidence** can be interpreted differently to support **opposite recommendations**

AI optimism

1. Black-box AI is **superior** to doctors;
2. Explainable AI is **useless**.
3. So we can **replace doctors with black-box AI** to improve medical decisions.

AI skepticism

1. This is a **small sample** study in an experimental setting. The result needs to be **further verified** in large-scale, real-world tasks.
2. Other factors can influence/explain away the results, such as **overreliance** on AI, and **weakness in XAI**.



Critical thinking requires us to have more prudence and holdback from claims like this one that may create significant impact.

Critical thinking & communication

The **same evidence** can be interpreted differently to support **opposite recommendations**

AI optimism

1. Black-box AI is **superior** to doctors;
2. Explainable AI is **useless**.
3. So we can **replace doctors with black-box AI** to improve medical decisions.



Critical thinking requires us to have more prudence and holdback from claims like this one that may create significant impact.

AI skepticism

1. This is a **small sample** study in an experimental setting. The result needs to be **further verified** in large-scale, real-world tasks.
2. Other factors can influence/explain away the results, such as **overreliance** on AI, and **weakness in XAI**.



That means we need to carefully consider the underlying assumptions and any possible counterarguments and counterevidence to them, especially from people with diverse perspectives and interests.

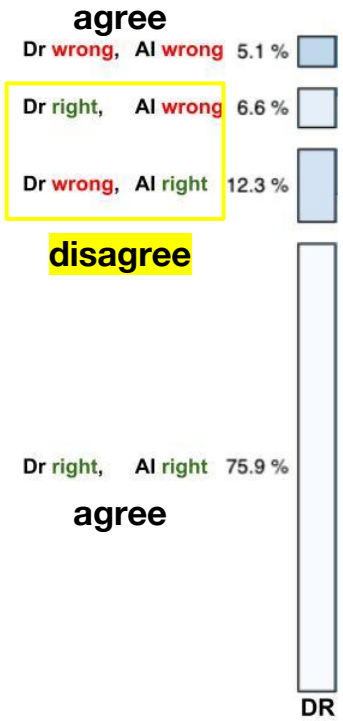
1. Why did doctors improve their performance with AI prediction?

For the total of 759 decisions (35 doctors × each read 25 images), we divided them into four groups:



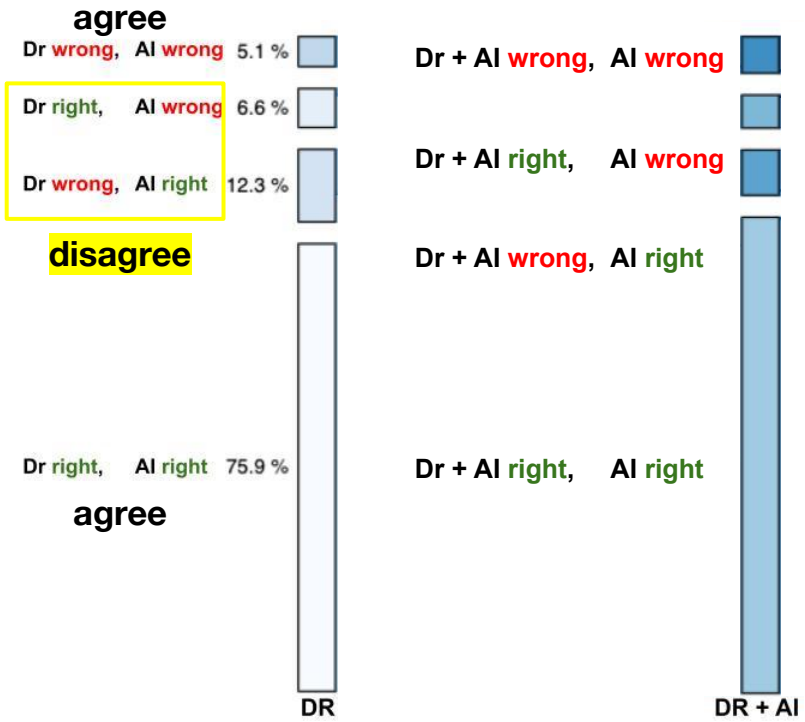
1. Why did doctors improve their performance with AI prediction?

For the total of 759 decisions (35 doctors × each read 25 images), we divided them into four groups:



1. Why did doctors improve their performance with AI prediction?

For the total of 759 decisions (35 doctors × each read 25 images), we divided them into four groups:

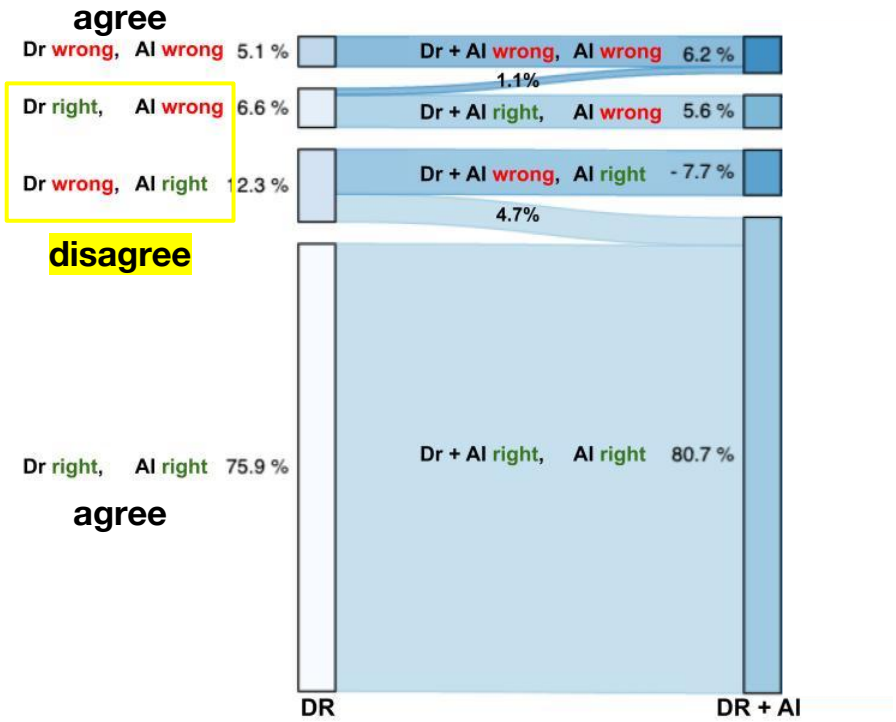


Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task. W Jin, M Fatehi, R Guo, G Hamarneh. AI in Medicine. 2024

1. Why did doctors improve their performance with AI prediction?

Because physicians' decision patterns converged to be more similar to AI decisions, as they **only switched decisions** when **disagreeing** with AI.

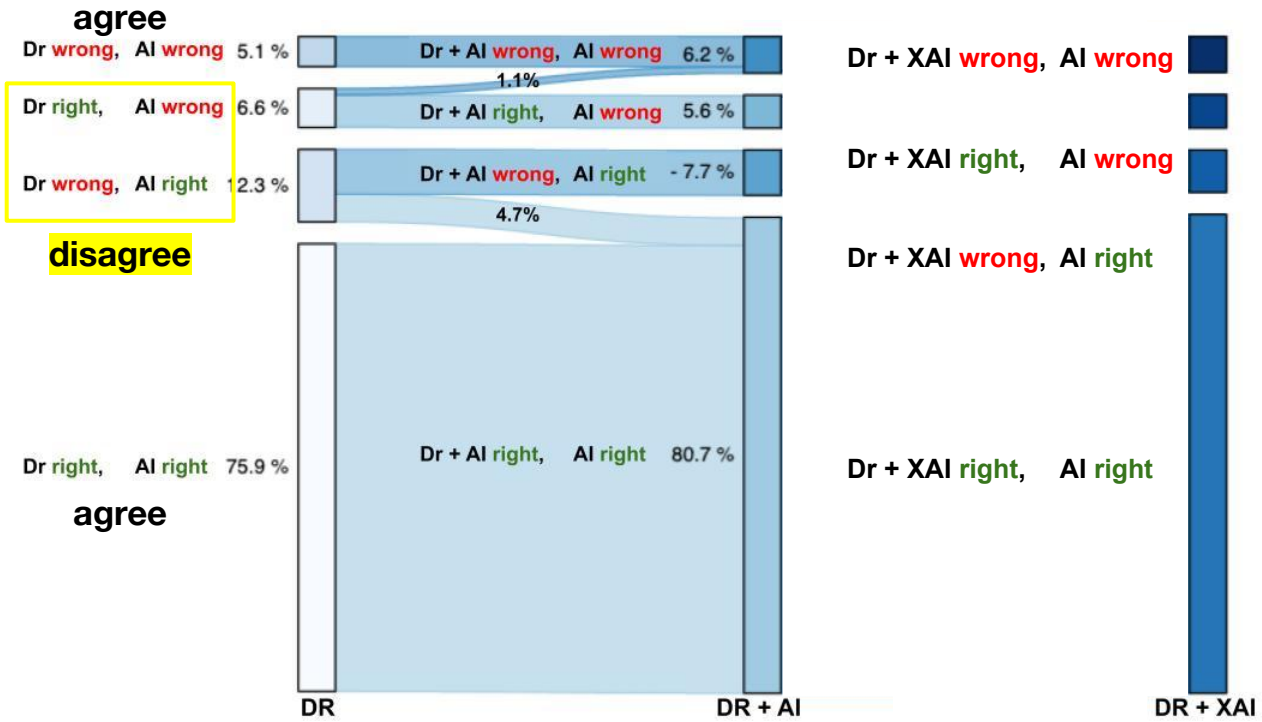
I.e.: doctors' improved performance is due to **overreliance on AI**



1. Why did doctors improve their performance with AI prediction?

Because physicians' decision patterns converged to be more similar to AI decisions, as they **only switched decisions** when **disagreeing** with AI.

I.e.: doctors' improved performance is due to **overreliance on AI**

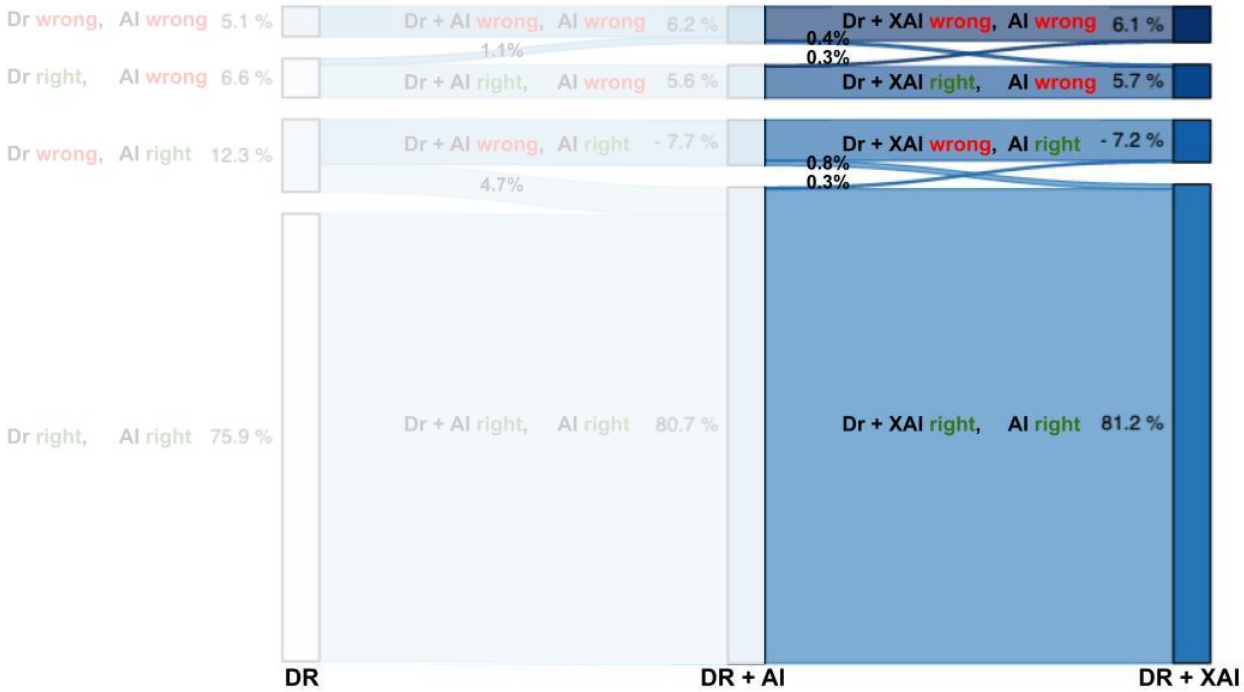


Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task. W Jin, M Fatehi, R Guo, G Hamarneh. AI in Medicine. 2024

2. Why was the AI explanation not helpful?

Because physicians had both **changed decisions *correctly*** **and *incorrectly*** with basically **the same amount**.

The **XAI algorithm is flawed**: Explanation cannot help doctors to discern potentially questionable decisions of AI.



Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task. W Jin, M Fatehi, R Guo, G Hamarneh. AI in Medicine. 2024

Critical thinking & communication

The same evidence can be interpreted differently to support opposite recommendations

- **AI optimism**
  Black-box AI is **superior** to doctors; explainable AI is **useless**. So we can **replace doctors with black-box AI** to improve medical decisions

- **AI skepticism**
  This is a **small sample** study in an experimental setting. The result needs to be **further verified** in large-scale, real-world tasks. Other factors can influence/explain away the results, such as **overreliance** on AI, and **weakness in XAI**



Critical thinking requires us to have more prudence and pushback from claims like this one that may create significant impact.



Take home message 2:
Because a piece of communication implies some values and perspectives, we need to carefully consider any possible counterarguments and counterevidence to it, especially from people with diverse perspectives and interests.

Critical communication

Take home message 1:
A piece of communication implies some goals, perspectives, values, and interests behind it, though some may appear to be scientific, objective, or value-neutral [1].

Take home message 2:
Because a piece of communication implies some values and perspectives, we need to carefully consider any possible counterarguments and counterevidence to it, especially from people with diverse perspectives and interests.

