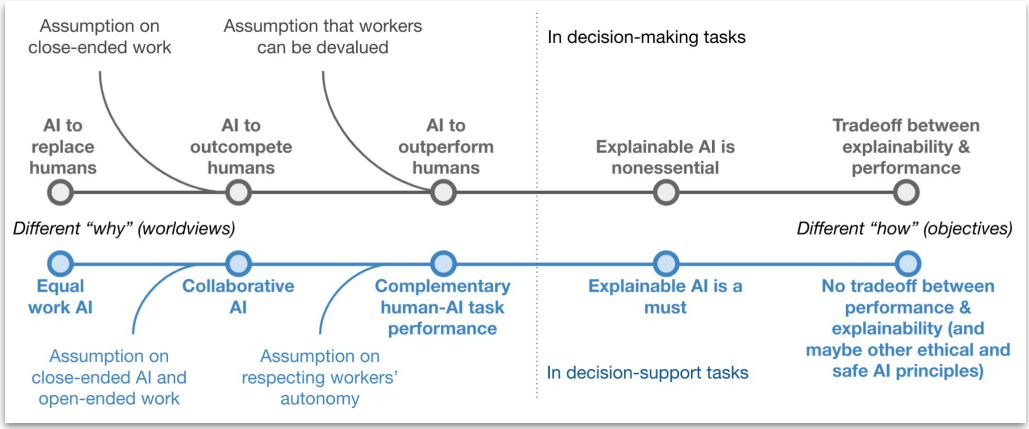
Constructing a Different Imagination Beyond "AI Outperforming Humans"

Weina Jin

Medical Image Analysis Lab

Advisors
Ghassan Hamarneh, Xiaoxiao Li



HAMARNEH
Medical Image Analysis Research Group

SFU AI

Why do we develop AI?

Why do we develop AI?

To free human labor by **outperforming and replacing humans**

To conquer difficult problems in society, such as cancer, climate change

To invent good tools that can extend human capability

To improve human science

AI is on trend and I do it for my career & money

To make profit for those who are in charge of AI tech

To prove human mind is a machine

That's the only thing I can do

To improve social welfare by raising **productivity** and **efficiency**

To upgrade human being to silicon-based life

To understand human being and intelligence by reverse engineering it

To accelerate human extinction

Why do we develop AI?

To free human labor by **outperforming and replacing humans**

To conquer difficult problems in society, such as cancer, climate change

To invent good tools that can extend human capability

To improve human science

AI is on trend and I do it for my career & money

To make profit for those who are in charge of AI tech

To prove human mind is a machine

That's the only thing I can do

To improve social welfare by raising **productivity** and **efficiency**

To upgrade human being to silicon-based life

To understand human being and intelligence by reverse engineering it

To accelerate human extinction

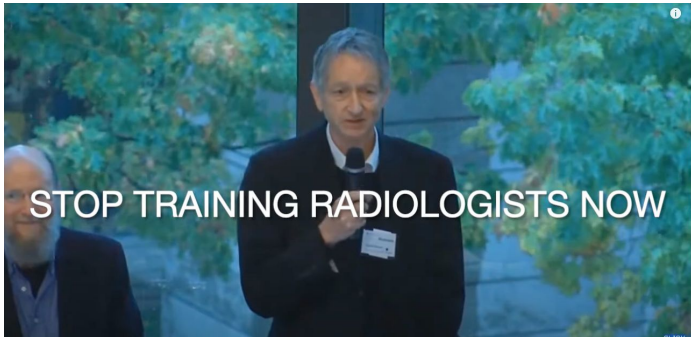
To free human labor by **outperforming and replacing humans**

Why do we develop AI?

The mainstream imagination of AI:
AI is to outperform and replace humans

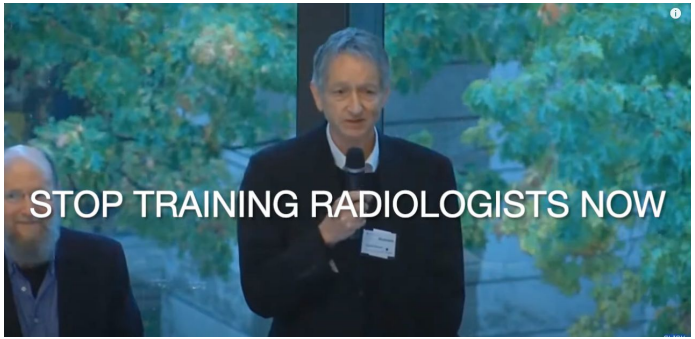
To improve social welfare by raising **productivity** and **efficiency**

The mainstream imagination of AI: to outperform and replace humans

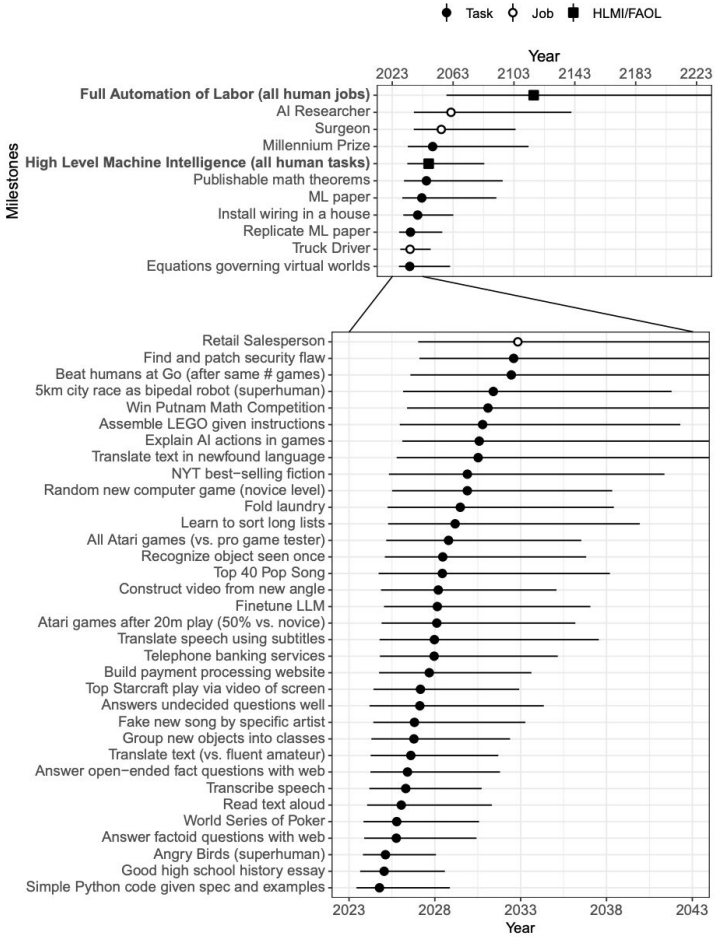


Geoffrey Hinton, 2016
https://www.youtube.com/watch?v=BxGyV2548xs

The mainstream imagination of AI: to outperform and replace humans



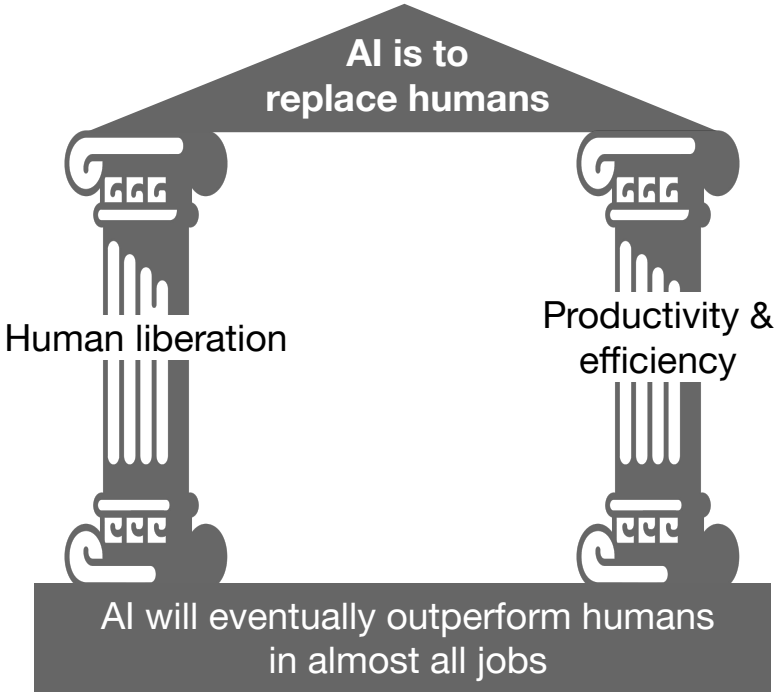
Geoffrey Hinton, 2016
https://www.youtube.com/watch?v=BxGyV2548xs



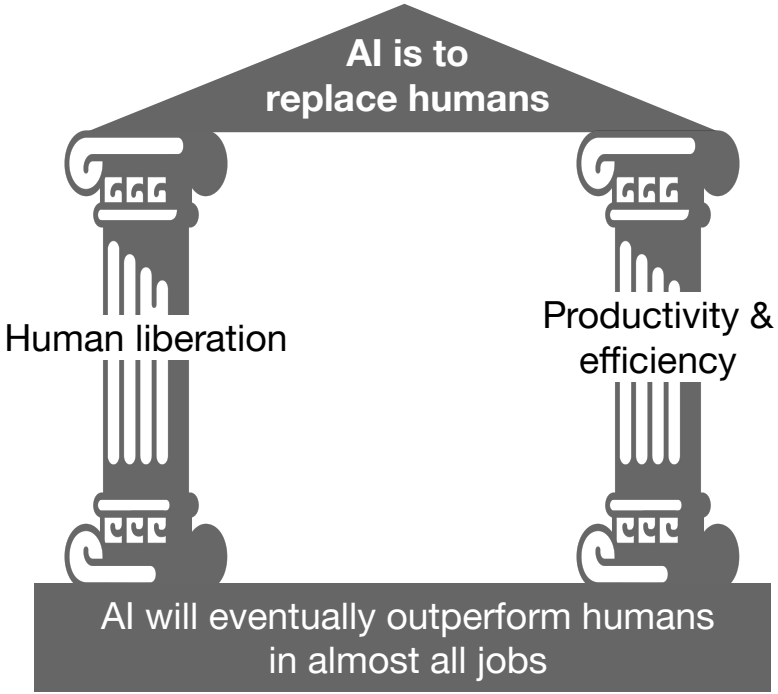
"High-level machine intelligence (HLMI) is achieved when **unaided machines can accomplish every task better and more cheaply than human workers**. Most milestones are predicted to have better than even odds of happening **within the next ten years**, though with a wide range of plausible dates. "

Thousands of AI Authors on the Future of AI. 2024. https://arxiv.org/pdf/2401.02843.pdf

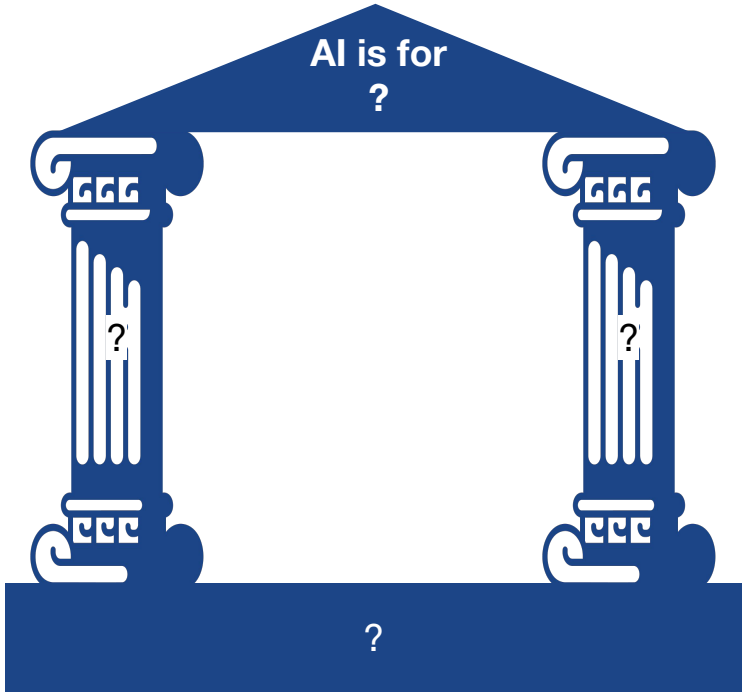
The mainstream imagination of AI



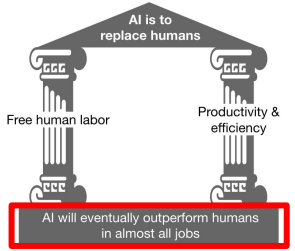
The mainstream imagination of AI



Imagination of AI for my thesis



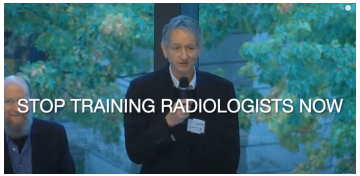
Argument against "AI will eventually surpass human-level performance in almost all jobs[1]"



[2]

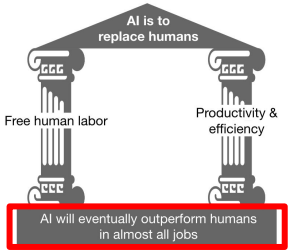
[3]

https://www.consumerreports.org/cars/autonomous-driving/no-you-cannot-buy-a-self-driving-car-today-a4355089516/



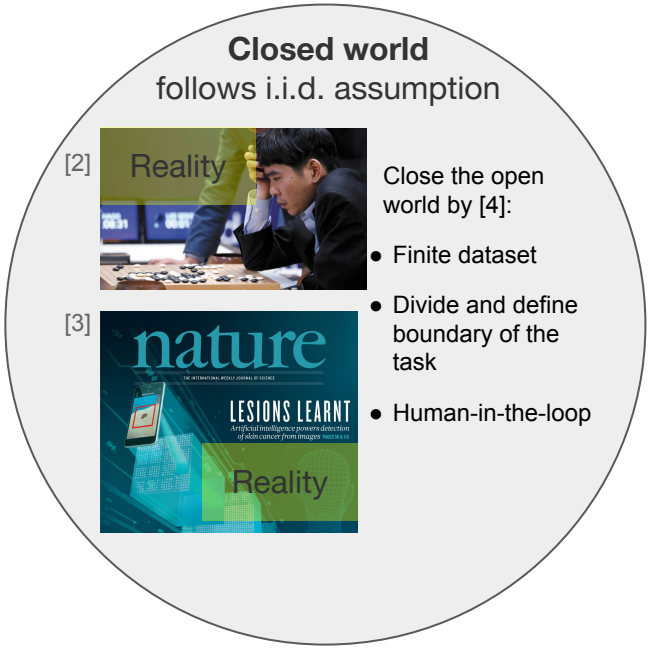
[1] Thousands of AI Authors on the Future of AI. 2024. https://arxiv.org/pdf/2401.02843.pdf
[2] Singh, S., Okun, A. & Jackson, A. Learning to play Go from scratch. Nature 550, 336–337 (2017).
[3] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017).
[4] Birhane, Abeba and David J. T. Sumpter. "The games we play: critical complexity improves machine learning." HHAI (2022).

Argument against "AI will eventually surpass human-level performance in almost all jobs[1]"



Open world

complex adaptive systems, e.g.: human minds, human behavior, financial market, cell, society



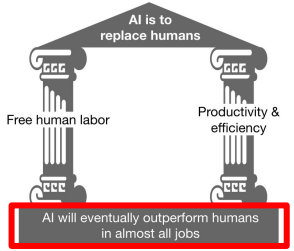
Critical complexity theory [4] and complexity science

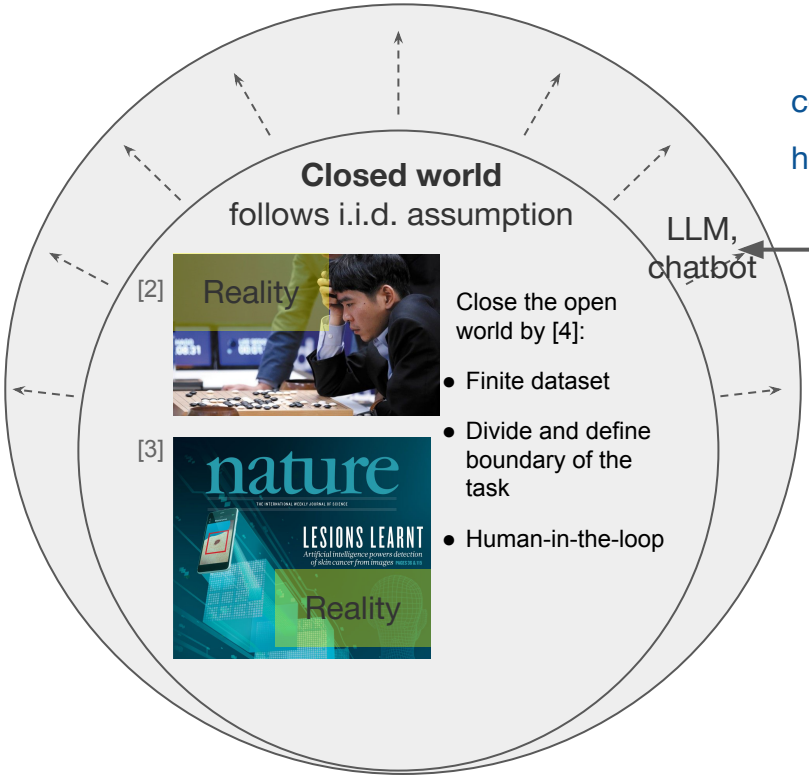
- No boundary of the system
- Irreducible, incompressible
- Non-linear, dynamic, stochastic interactions give rise to emergent behaviour
- Historical, contextual, value-laden
- Impossible to model "**All models are wrong**, some are useful."

https://www.consumerreports.org/cars/autonomous-driving/no-you-cannot-buy-a-self-driving-car-today-a4355089516/

[1] Thousands of AI Authors on the Future of AI. 2024. https://arxiv.org/pdf/2401.02843.pdf
[2] Singh, S., Okun, A. & Jackson, A. Learning to play Go from scratch. Nature 550, 336–337 (2017).
[3] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017).
[4] Birhane, Abeba and David J. T. Sumpter. "The games we play: critical complexity improves machine learning." HHAI (2022).

Argument against "AI will eventually surpass human-level performance in almost all jobs[1]"





Open world

complex adaptive systems, e.g.: human minds, human behavior, financial market, cell, society

LLM, chatbot ← language, customer service

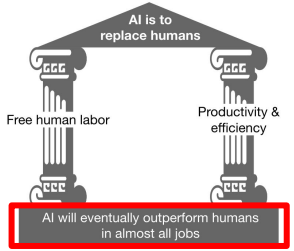
Critical complexity theory [4] and complexity science

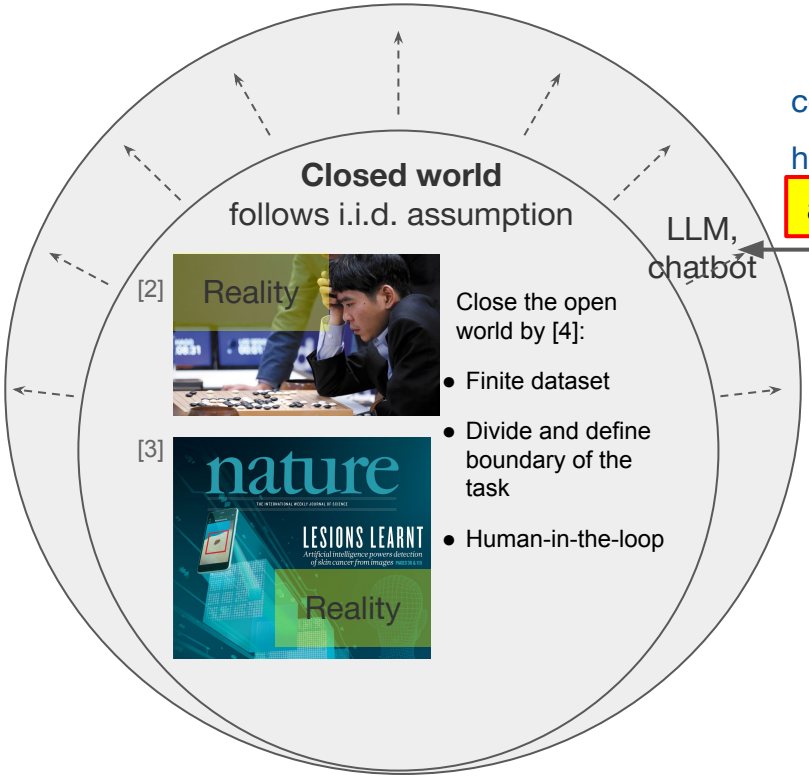
- No boundary of the system
- Irreducible, incompressible
- Non-linear, dynamic, stochastic interactions give rise to emergent behaviour
- Historical, contextual, value-laden
- Impossible to model

https://www.consumerreports.org/cars/autonomous-driving/no-you-cannot-buy-a-self-driving-car-today-a4355089516/

[1] Thousands of AI Authors on the Future of AI. 2024. https://arxiv.org/pdf/2401.02843.pdf
[2] Singh, S., Okun, A. & Jackson, A. Learning to play Go from scratch. Nature 550, 336–337 (2017).
[3] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017).
[4] Birhane, Abeba and David J. T. Sumpter. "The games we play: critical complexity improves machine learning." HHAI (2022).

Argument against "AI will eventually surpass human-level performance in almost all jobs[1]"





Open world

complex adaptive systems, e.g.: human minds, human behavior, financial market, cell, society

at cost

language, customer service

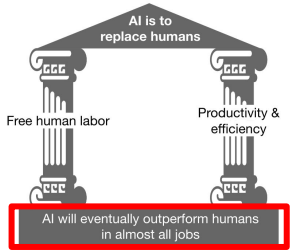
Critical complexity theory [4] and complexity science

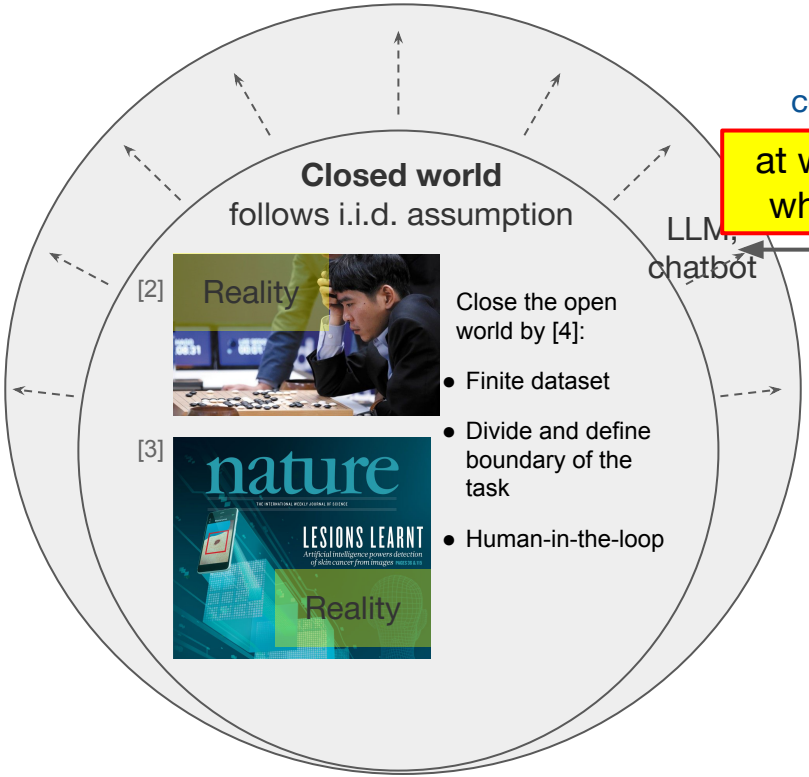
- No boundary of the system
- Irreducible, incompressible
- Non-linear, dynamic, stochastic interactions give rise to emergent behaviour
- Historical, contextual, value-laden
- Impossible to model

https://www.consumerreports.org/cars/autonomous-driving/no-you-cannot-buy-a-self-driving-car-today-a4355089516/

[1] Thousands of AI Authors on the Future of AI. 2024. https://arxiv.org/pdf/2401.02843.pdf
[2] Singh, S., Okun, A. & Jackson, A. Learning to play Go from scratch. Nature 550, 336–337 (2017).
[3] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017).
[4] Birhane, Abeba and David J. T. Sumpter. "The games we play: critical complexity improves machine learning." HHAI (2022).

Argument against "AI will eventually surpass human-level performance in almost all jobs[1]"





Open world

complex adaptive systems, e.g.: human minds,

financial market, cell, society

at what cost & whose cost?

language, customer service

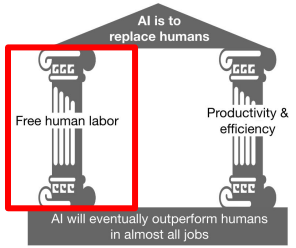
Critical complexity theory [4] and complexity science

- No boundary of the system
- Irreducible, incompressible
- Non-linear, dynamic, stochastic interactions give rise to emergent behaviour
- Historical, contextual, value-laden
- Impossible to model

https://www.consumerreports.org/cars/autonomous-driving/no-you-cannot-buy-a-self-driving-car-today-a4355089516/

[1] Thousands of AI Authors on the Future of AI. 2024. https://arxiv.org/pdf/2401.02843.pdf
[2] Singh, S., Okun, A. & Jackson, A. Learning to play Go from scratch. Nature 550, 336–337 (2017).
[3] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017).
[4] Birhane, Abeba and David J. T. Sumpter. "The games we play: critical complexity improves machine learning." HHAI (2022).

Replacing human labor **with** AI **will free** humans

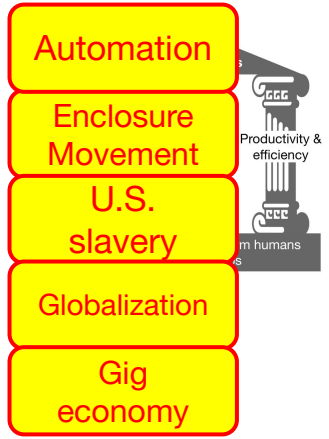


[1] Neda Atanasoski; Kalindi Vora. Surrogate Humanity: Race, Robots, and the Politics of Technological Futures. Duke University Press. 2019
[2] Mary L. Gray and Siddharth Suri. Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass

	human labor		AI		humans	Automation
	peasants		women			Enclosure Movement
Replacing	peasants	**with**	slaves	**will free**	the wealthy and powerful	U.S. slavery
	local workers		offshore workers			Globalization
	full-time workers		gig workers			Gig economy

[1] Neda Atanasoski; Kalindi Vora. Surrogate Humanity: Race, Robots, and the Politics of Technological Futures. Duke University Press. 2019
[2] Mary L. Gray and Siddharth Suri. Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass

	human labor		AI		humans
	peasants		women		the wealthy and powerful
Replacing	peasants	with	slaves	will free	
	local workers		offshore workers		
	full-time workers		gig workers		



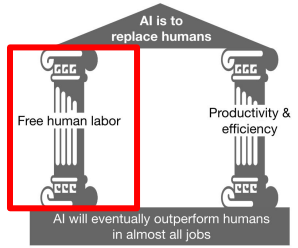
- **cheaper** and **do not rebel**
- some workers and their work are **devalued** or **degraded** due to their social hierarchies, such that they can be atomized into a relatively closed form, harder to unionize
- assumes a new human-machine **hierarchy** that reproduces the logic of racial, ethnic, and gender hierarchies [1]

[1] Neda Atanasoski; Kalindi Vora. Surrogate Humanity: Race, Robots, and the Politics of Technological Futures. Duke University Press. 2019
[2] Mary L. Gray and Siddharth Suri. Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass

Replacing	human labor	with	AI	will free
	peasants		women	
	peasants		slaves	
	local workers		offshore workers	
	full-time workers		gig workers	

~~humans~~

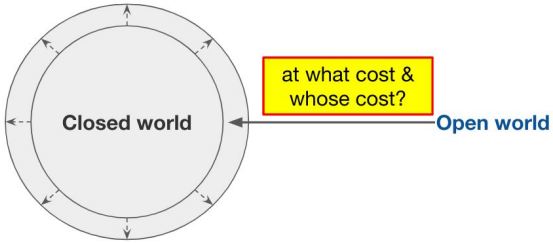
the wealthy and powerful



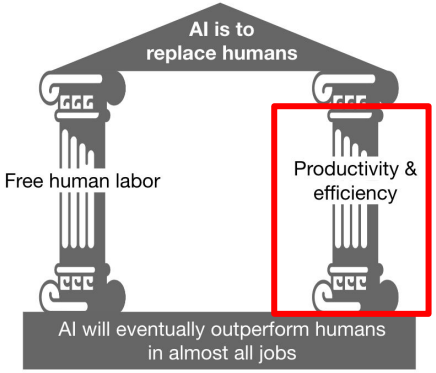
- **cheaper** and **do not rebel**
- some workers and their work are **devalued** or **degraded** due to their social hierarchies, such that they can be atomized into a relatively closed form, harder to unionize
- assumes a new human-machine **hierarchy** that reproduces the logic of racial, ethnic, and gender hierarchies [1]

Who and whose work are devalued or degraded when AI replaces humans?

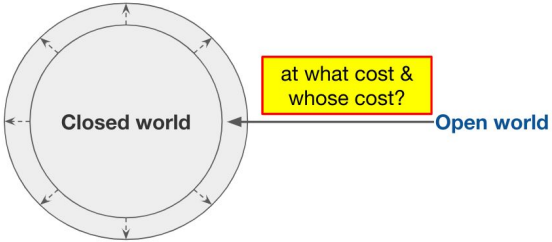
- Workers who are replaced by AI
- Workers who work behind AI [2]



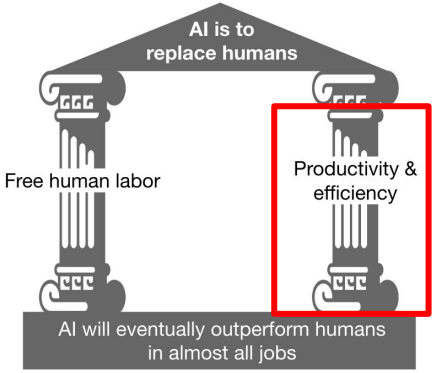
[1] Neda Atanasoski; Kalindi Vora. Surrogate Humanity: Race, Robots, and the Politics of Technological Futures. Duke University Press. 2019
[2] Mary L. Gray and Siddharth Suri. Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass



Replacing	human labor	with	AI	Worker replacement to increase productivity, **for whom?**
	peasants		women	
	peasants		slaves	
	local workers		offshore workers	
	full-time workers		gig workers	



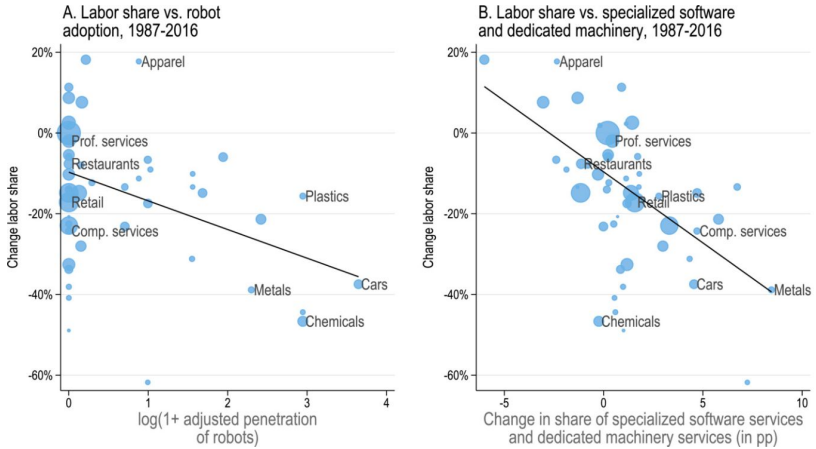
[1] Acemoglu, D. and Restrepo, P. (2022), Tasks, Automation, and the Rise in U.S. Wage Inequality. Econometrica, 90: 1973-2016. https://doi.org/10.3982/ECTA19815

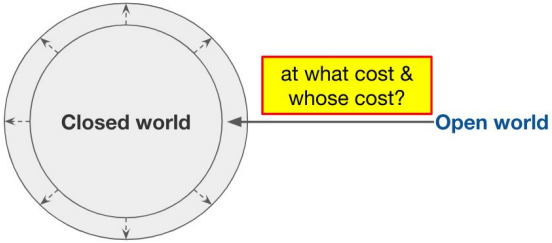


Replacing	human labor	with	AI
	peasants		women
	peasants		slaves
	local workers		offshore workers
	full-time workers		gig workers

Worker replacement to increase productivity,

for whom?

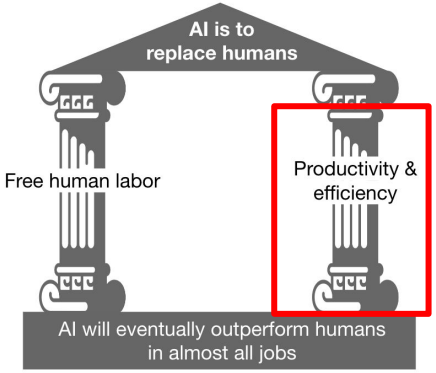




Automation and productivity gain can go hand-in-hand with widened inequality [1]

Negative correlation between automation technologies and changes in industry labor shares [1].

[1] Acemoglu, D. and Restrepo, P. (2022), Tasks, Automation, and the Rise in U.S. Wage Inequality. Econometrica, 90: 1973-2016. https://doi.org/10.3982/ECTA19815



Replacing		with	
	human labor		AI
	peasants		women
	peasants		slaves
	local workers		offshore workers
	full-time workers		gig workers

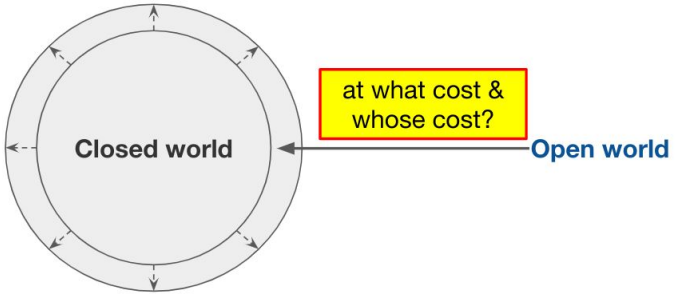
Worker replacement to increase productivity,

for whom?

at what cost and whose cost?

At the cost of other essential human and social values [1,2,3]**:**

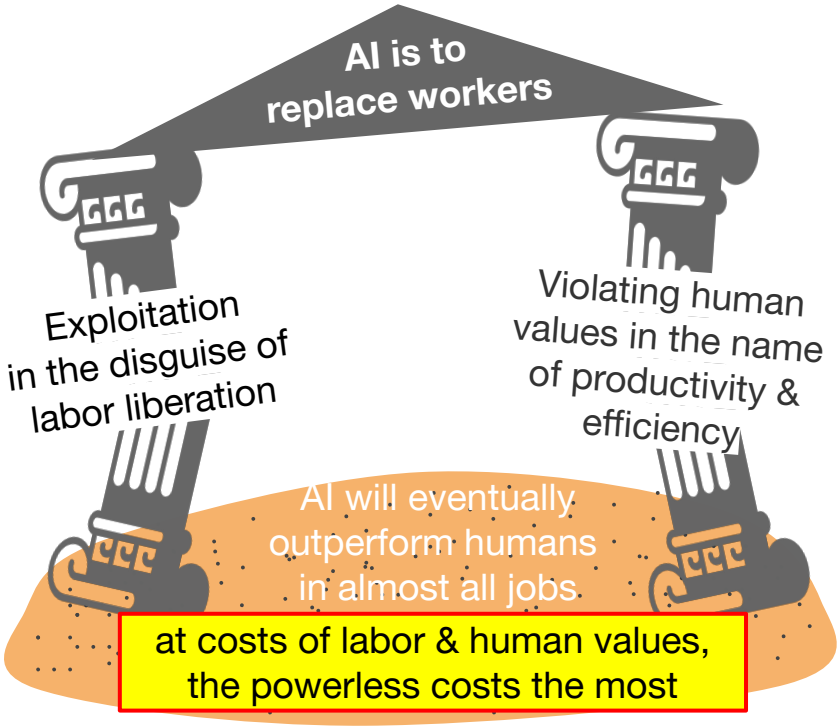
- Human rights, autonomy, equity, social justice, democracy, privacy, copyright, the value of work, environmental values, user experience, community, belonging……



The costs and harms are disproportionally distributed among the powerful and the powerless [1,2,3]

[1] Chris Wiggins and Matthew L. Jones. How Data Happened: A History from the Age of Reason to the Age of Algorithms. WW Norton. 2023
[2] Kate Crawford. Atlas of AI Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press. 2021
[3] Cathy O'Neil. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Books. 2016

The mainstream imagination of AI

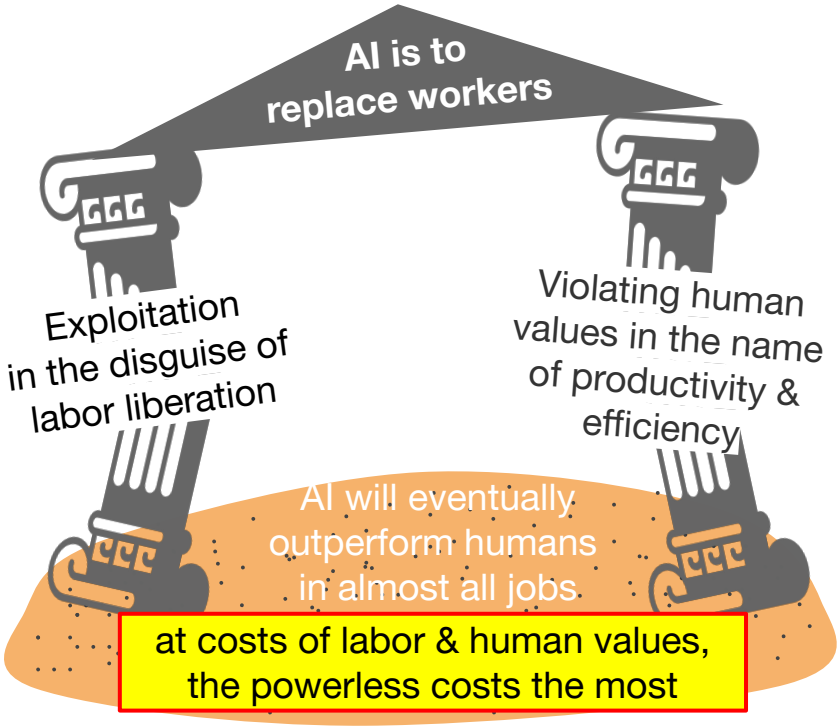


"People who criticize new technologies are sometimes called ***Luddites****, but it's helpful to clarify what the Luddites actually wanted. The main thing they were protesting was the fact that* ***their wages were falling*** *at the same time that* ***factory owners' profits were increasing****,* ***along with food prices****. They were also* ***protesting unsafe working conditions****,* ***the use of child labor****, and* ***the sale of shoddy goods*** *that discredited the entire textile industry. The Luddites did not indiscriminately destroy machines; if a machine's owner paid his workers well, they left it alone. The Luddites were* ***not anti-technology****; what they wanted was* ***economic justice****. They destroyed machinery as a way to get factory owners' attention. The fact that the word "Luddite" is now used as an insult, a way of calling someone irrational and ignorant, is a result of* ***a smear campaign by the forces of capital****."*

— Ted Chiang, *"Will A.I. Become the New McKinsey?"*

[1] Chris Wiggins and Matthew L. Jones. How Data Happened: A History from the Age of Reason to the Age of Algorithms. WW Norton. 2023

The mainstream imagination of AI



Technology and its impact are the symptom (“fever”),

not the root cause (“disease”).

The treatment effect will be limited if we just blame and correct the bad datasets/algorithms/technologists (“lower the temperature”)

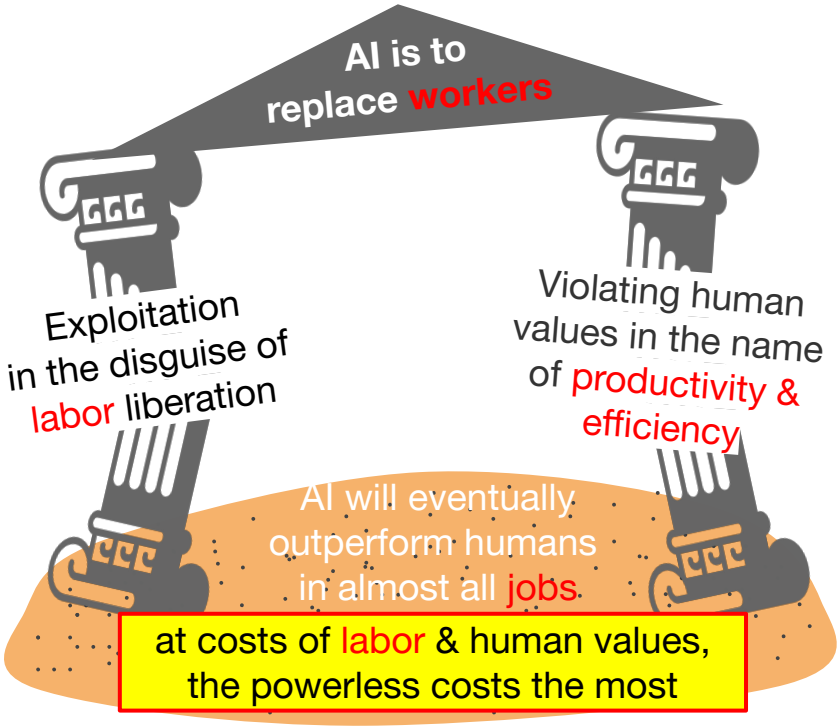
without identifying and treating the underlying disease. [1]

“The Luddites did not indiscriminately destroy machines; if a machine’s owner paid his workers well, they left it alone. The Luddites were ***not anti-technology****; what they wanted was* ***economic justice****.”*

— Ted Chiang, *“Will A.I. Become the New McKinsey?”*

[1] Chris Wiggins and Matthew L. Jones. How Data Happened: A History from the Age of Reason to the Age of Algorithms. WW Norton. 2023

The mainstream imagination of AI



Technology and its impact are the symptom (“fever”),

not the root cause (“disease”).

The treatment effect will be limited if we just blame and correct the bad datasets/algorithms/technologists (“lower the temperature”)

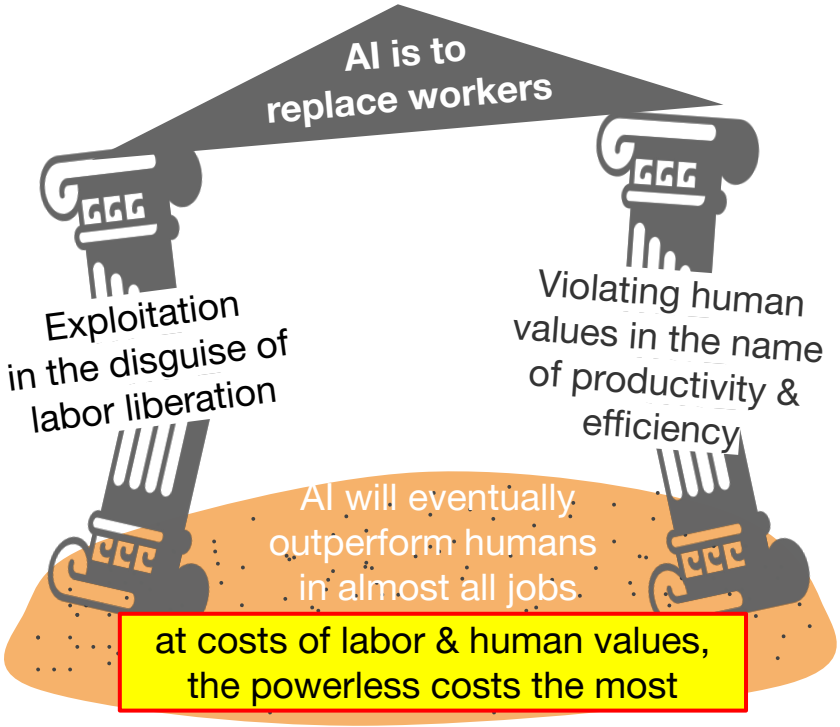
without identifying and treating the underlying disease. [1]

My diagnosis of the root cause (“disease”):

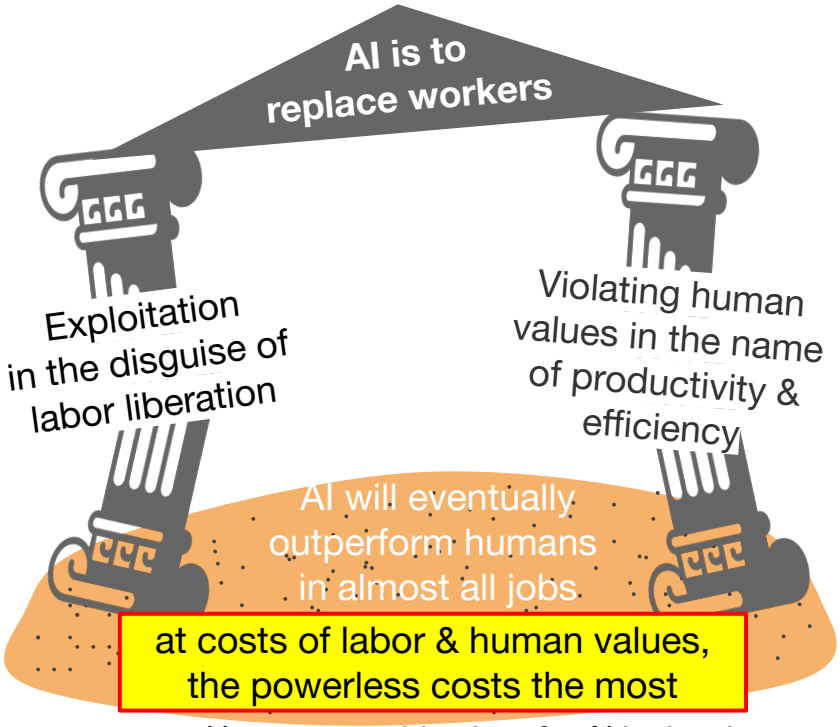
AI is deeply entangled with the concept of work. I.e.: the means of production and how a society and its social, political, and economic structures are organized around it.

[1] Chris Wiggins and Matthew L. Jones. How Data Happened: A History from the Age of Reason to the Age of Algorithms. WW Norton. 2023

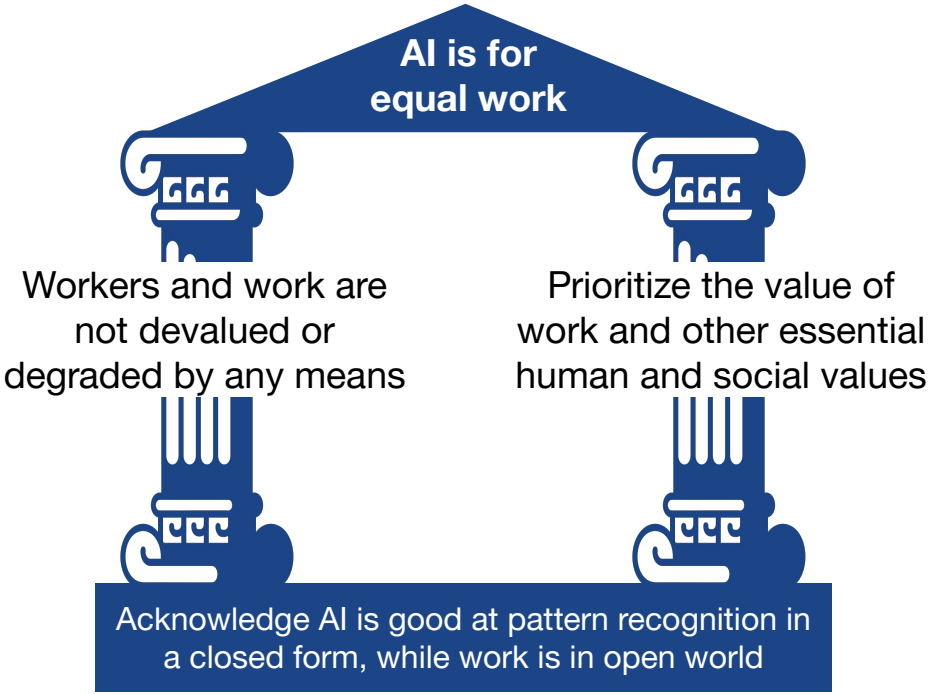
The mainstream imagination of AI



The mainstream imagination of AI



Imagination of AI for my project

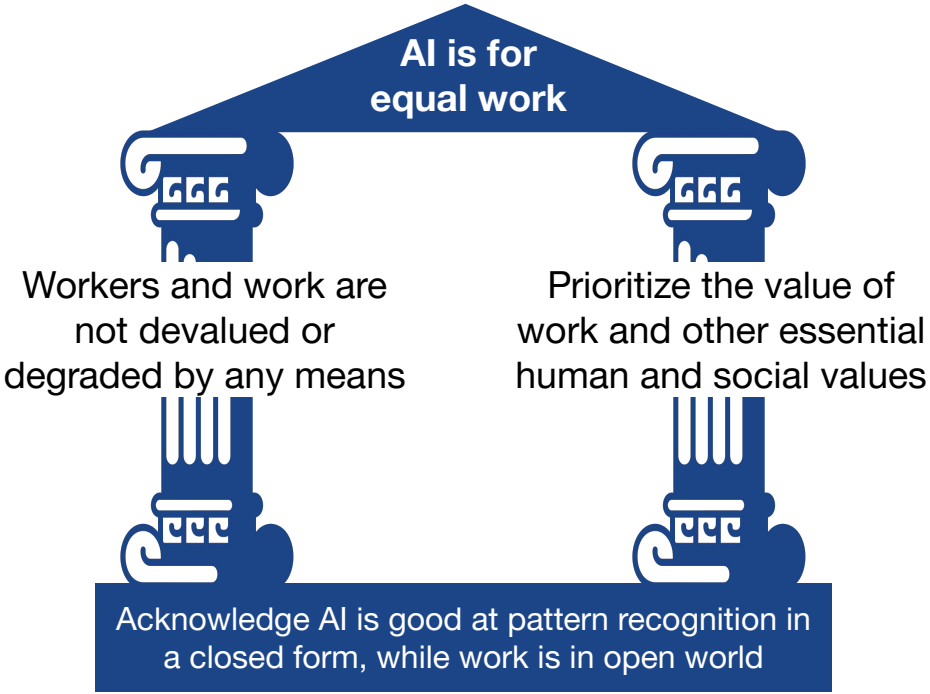


How we set objectives for AI is deeply entangled with different values and moralities for **work**, reflecting different **ideologies** in society

View **work** as laborious that should be avoided.
Value the freedom in consumption over the freedom in production

View **work** as an indispensable human activity that create value, meaning, belonging, and connection in community.
Value the freedom in production over the freedom in consumption

Imagination of AI for my project



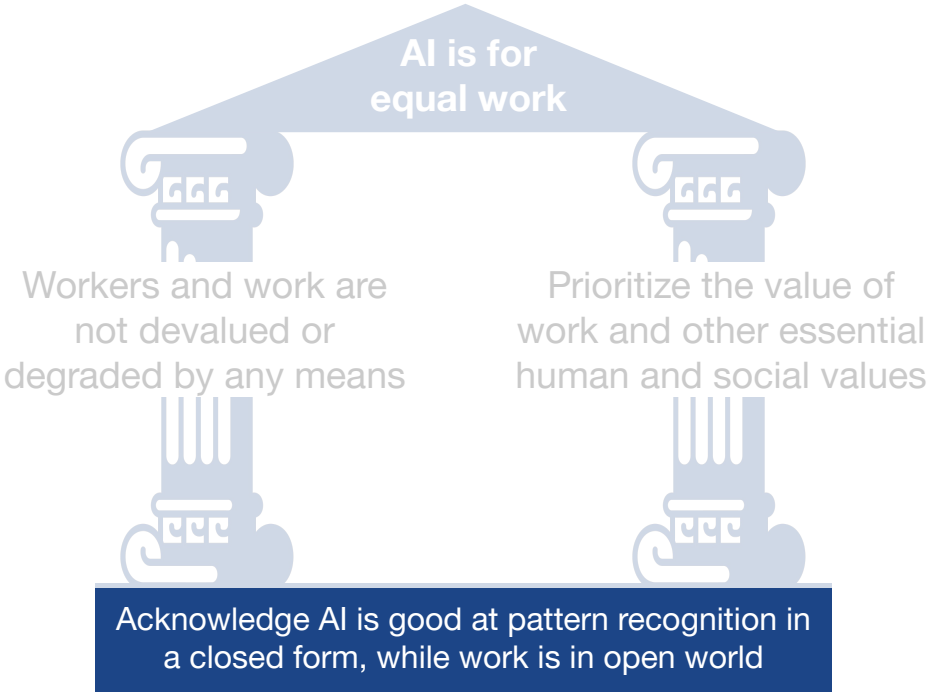
View **work** as an indispensable human activity that create value, meaning, belonging, and connection in community.
Value the freedom in production over the freedom in consumption

What will AI be under this new "axiomatic system" (worldview)?

I use **medical image analysis** as the real-world task domain

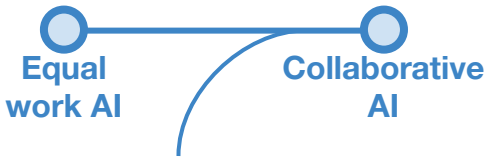
Equal work AI

Imagination of AI for my project



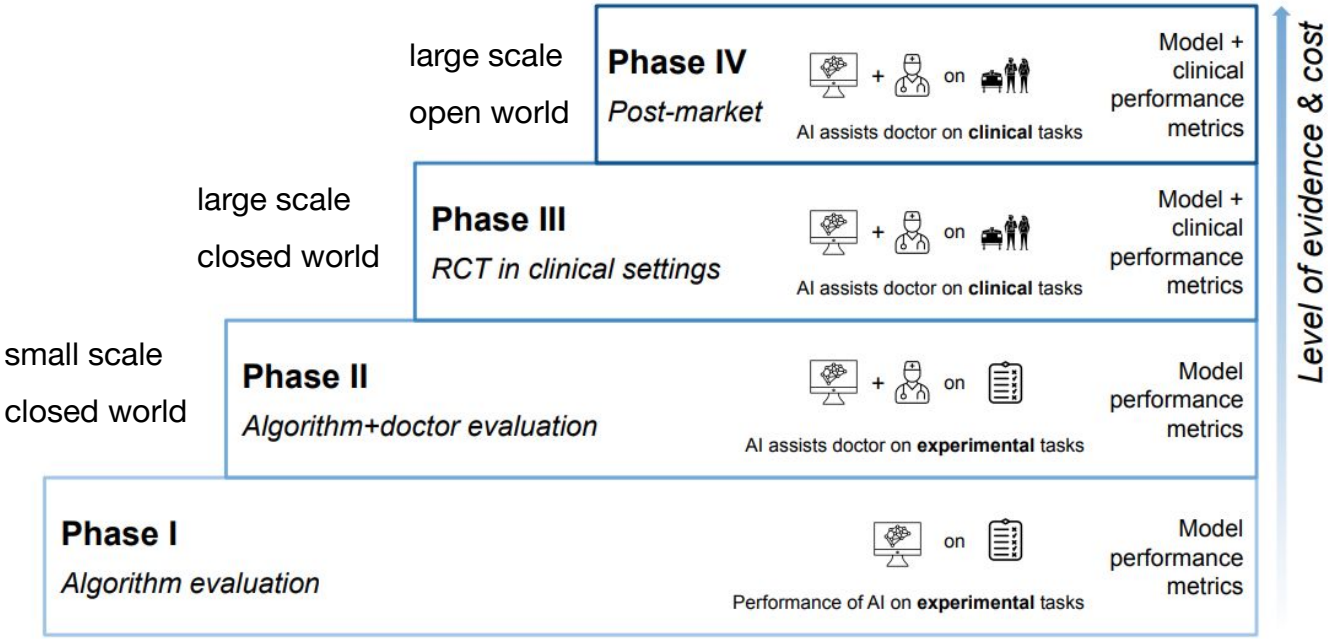
What will AI be under this new "axiomatic system" (worldview)?

I use **medical image analysis** as the real-world task domain



Assumption on close-ended AI and open-ended task

Testing AI for human-AI collaboration in real-world clinical settings

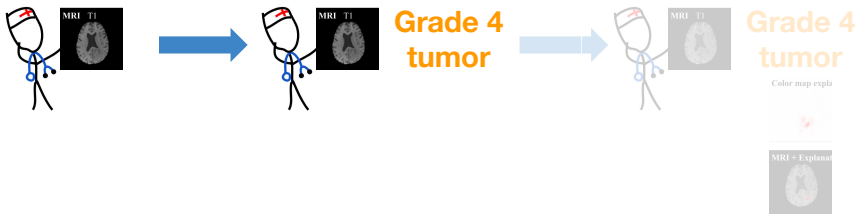


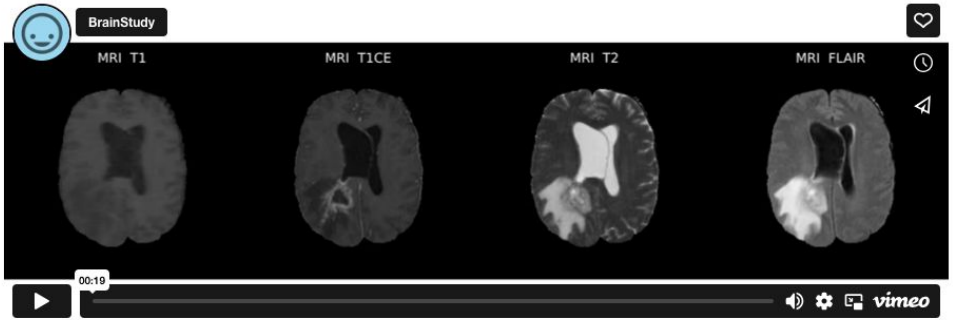
The four phases of evaluating the clinical utility of AI in glioma imaging [1].

[1] Weina Jin, Mostafa Fatehi, Kumar Abhishek, Mayur Mallya, Brian Toyota, Ghassan Hamarneh. Artificial intelligence in glioma imaging: challenges and advances. J Neural Eng. 2020

A clinical user study on the utility of AI in doctor-AI collaboration

- National online survey, 35 neurosurgeons,
  each read 25 MRI, gave judgment at three conditions:





1. **DR**
Doctor only

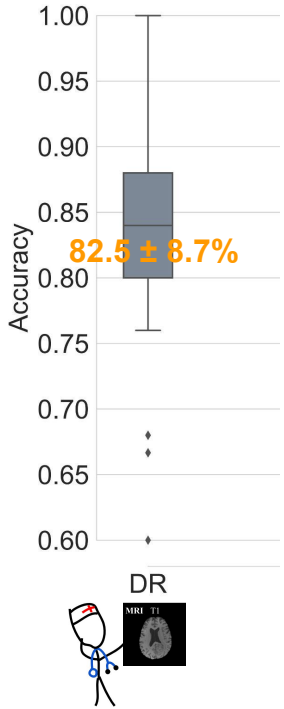
2. **DR+AI**
Doctor assisted by AI **prediction**

3. **DR+XAI**
Doctor assisted by AI prediction & **explanation**

Study Ethics No.: H20-03588
Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

Result

Is doctor+AI better than doctor alone?



Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

Result

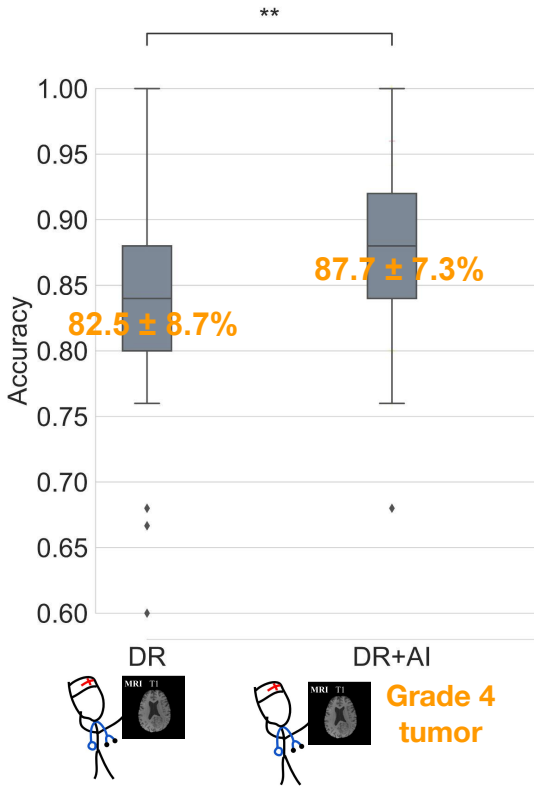
Is doctor+AI better than doctor alone?

Wilcoxon signed-rank tests with Bonferroni correction

Doctor + AI > Doctor ✅

On improving doctors' task performance:

- AI prediction (**DR+AI**) is helpful



Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

Result

Is doctor+AI better than doctor only?

Wilcoxon signed-rank tests with Bonferroni correction

Doctor + AI > Doctor ✅

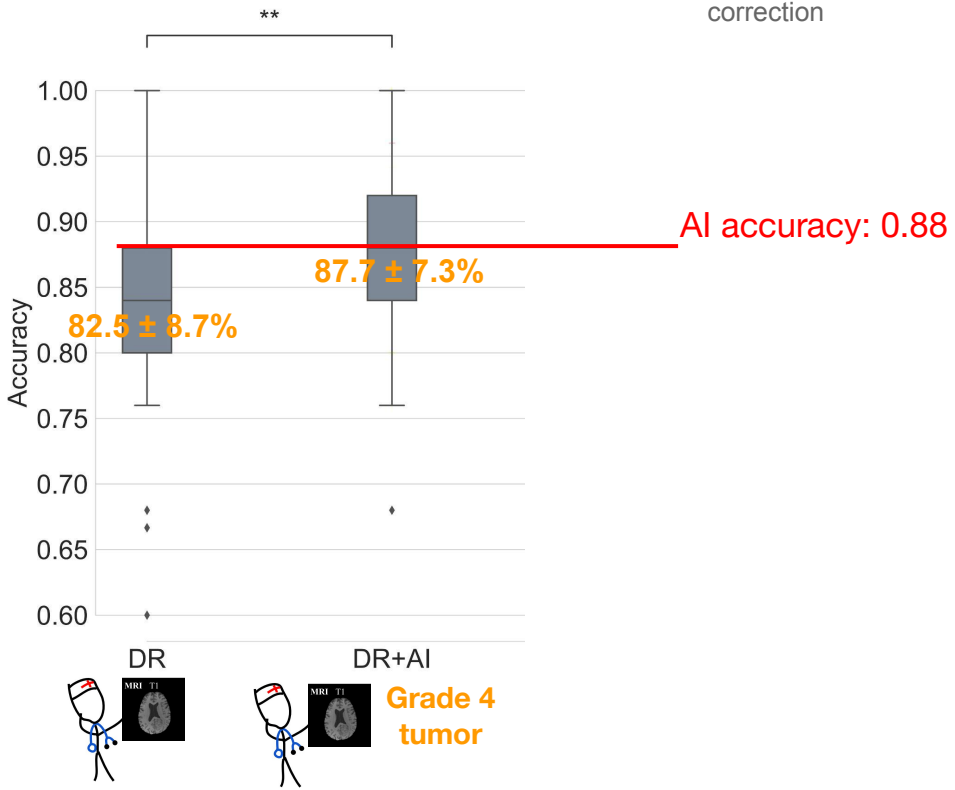
On improving doctors' task performance:

- AI prediction (**DR+AI**) is helpful

If so,

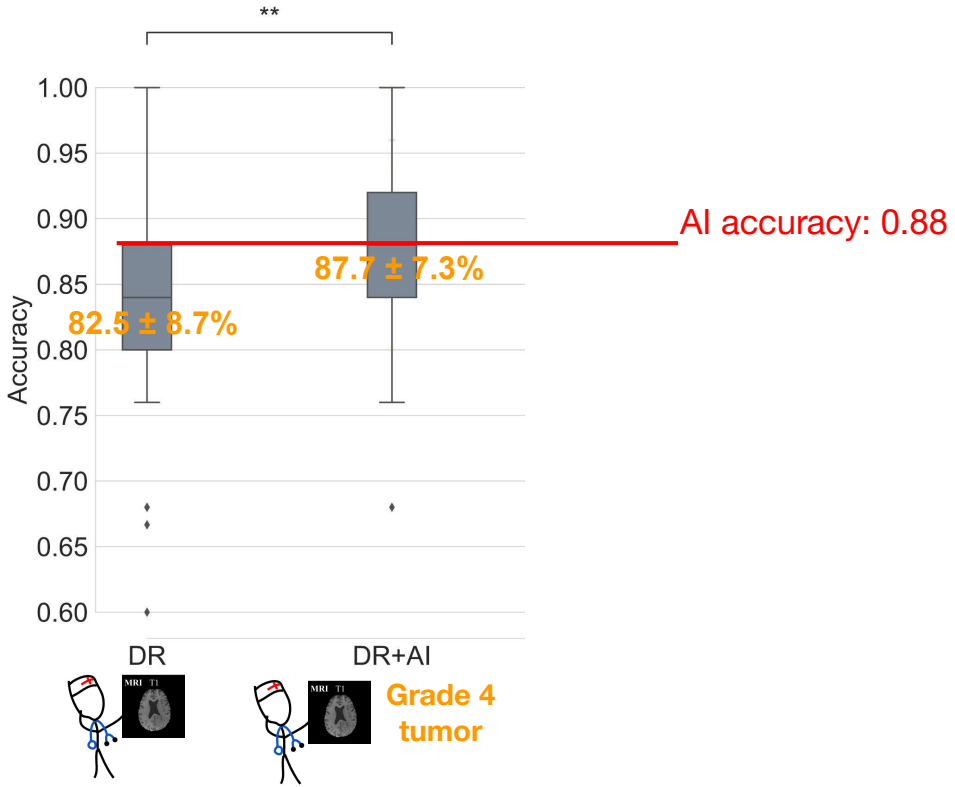
Doctor + AI > max(Doctor, AI) ❌

Not achieve complementary doctor-AI performance in **DR+AI**



Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

Critical moment Should we replace doctors with AI? Why or why not?



Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

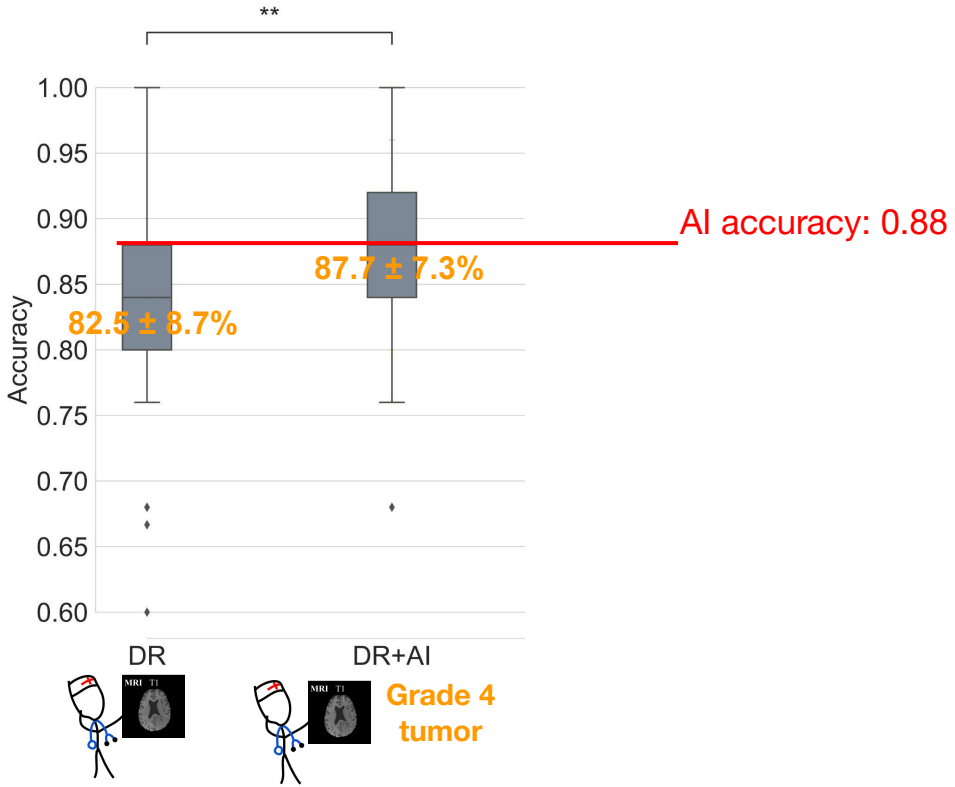
Critical moment Should we replace doctors with AI? Why or why not?

AI to replace humans

- Ignores the fact that evaluation is conducted in a closed form; devalue worker's value in open-world tasks
- Regard machine as rational, perfect. And humans as irrational, imperfect. Adjust humans to AI by degrading worker's work and closing the open world task

Equal work AI

- Adjust AI to humans

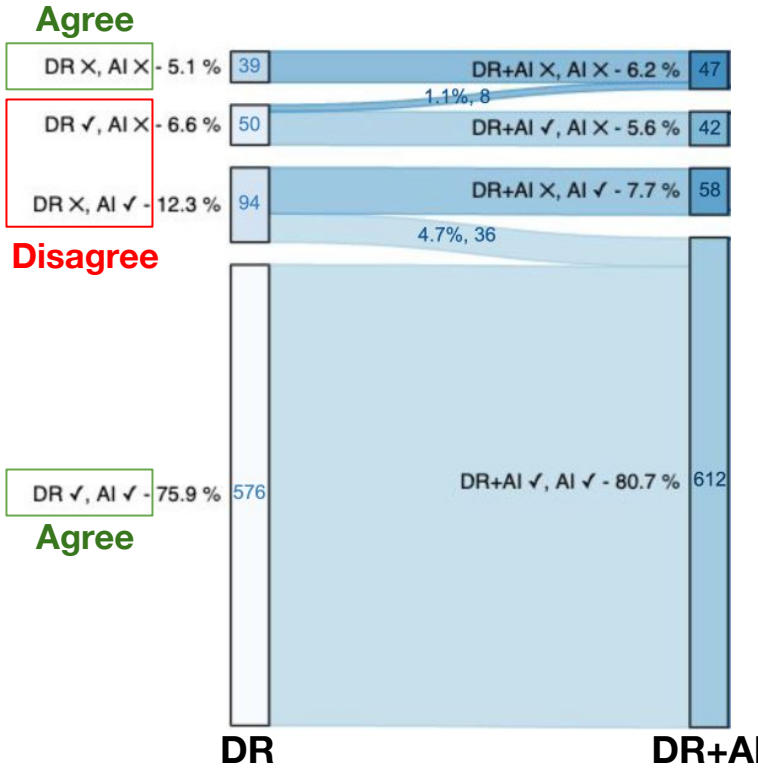


Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

Result Why doctors improved 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**

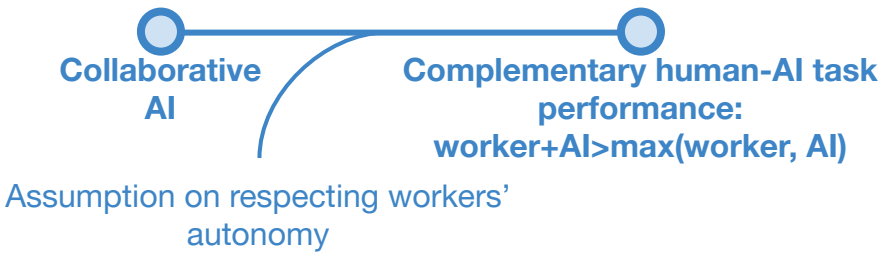


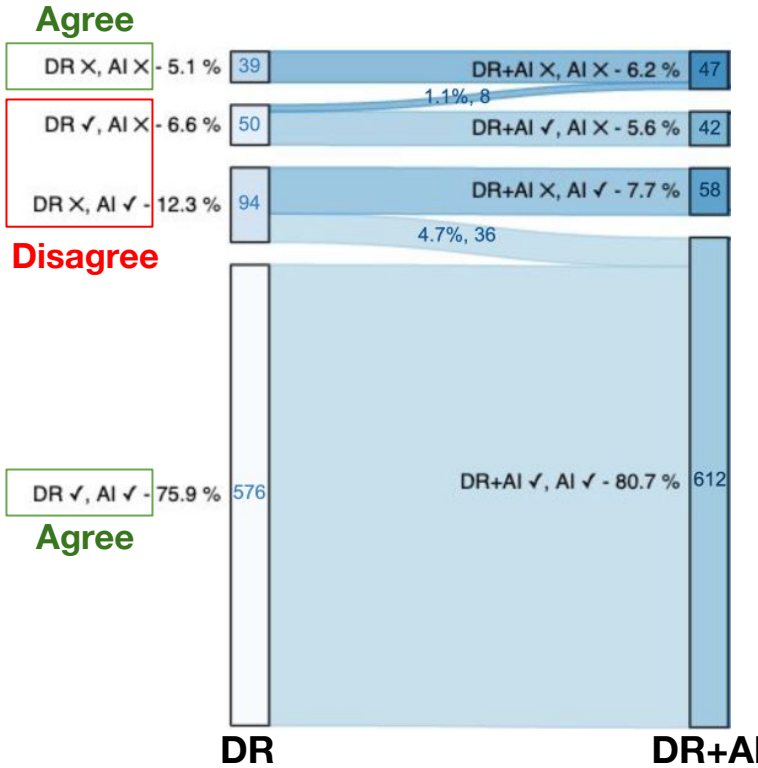
Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

Result Why doctors improved 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**

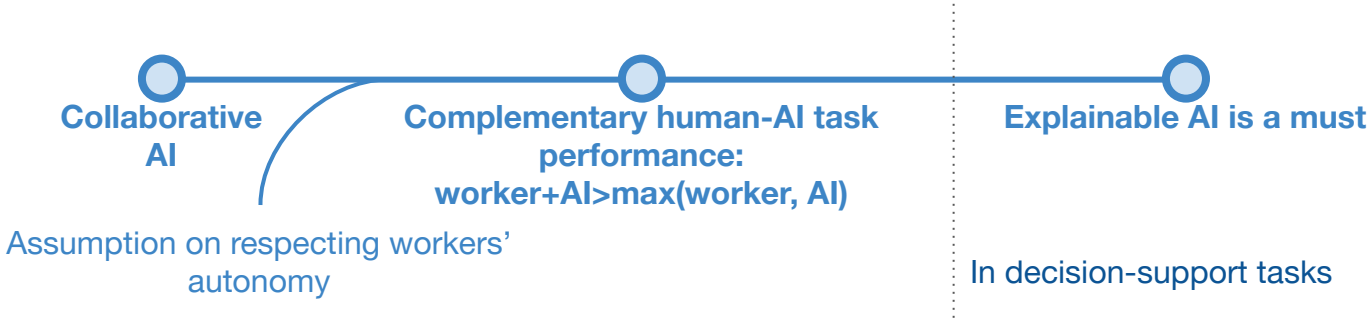




Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

How to achieve complementary human-AI task performance?

Theorem 1. *With a black-box AI as the human's decision-making assistant, and the task accuracies h and m of the human and AI, then the human-AI team accuracy t will not achieve complementary accuracy, i.e.:* $t \leq \max(h, m)$.



Proof. Since the black-box AI will provide the human with information that is not instance-specific, the human has a uniform probability of whether to take or reject an AI suggestion, we denote such uniform probability as b. Then the human-AI team accuracy can be calculated as:

$$t = mb + h(1-b)$$

If $h \geq m$:

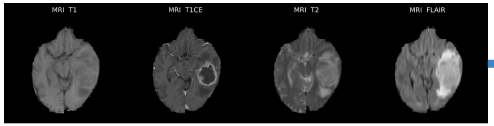
$$t = mb + h(1-b) \leq hb + h - hb = h$$

If $h < m$:

$$t = mb + h(1-b) < mb + m - mb = m$$

Therefore, $t \leq \max(h, m)$

Explainable AI (XAI) the additional information that facilitate workers' proper reliance on AI prediction

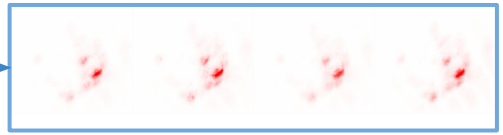


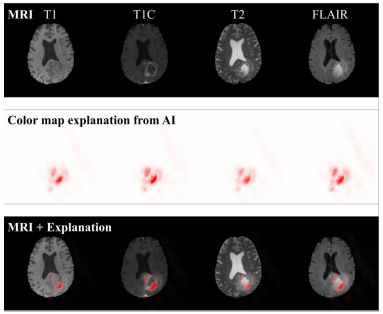
Input MRI

3D CNN model

AI's suggestion:
Grade 4 glioblastoma

AI's explanation:





Complementary human-AI task performance

Explainable AI is a must

In decision-support tasks

Candidates of 16 post-hoc heatmap explanation methods on the glioma task

Gradient based

Grad-CAM

Gradient

Input x Gradient

SmoothGrad

Deconvolution

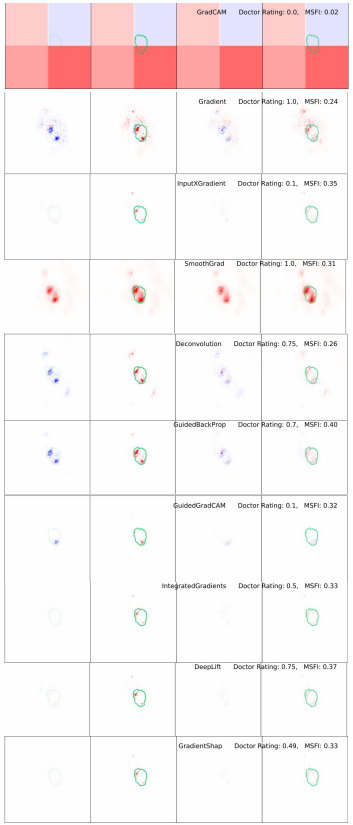
Guided Backpropagation

Guided Grad-CAM

Integrated Gradient

DeepLIFT

Gradient SHAP



Perturbation based

Occlusion

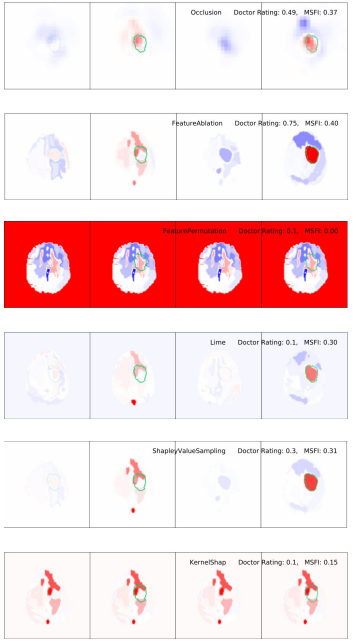
Feature Ablation

Feature Permutation

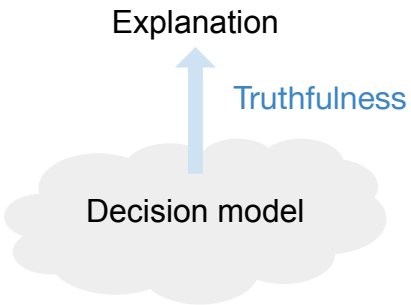
LIME

Shapley Value Sampling

Kernel SHAP

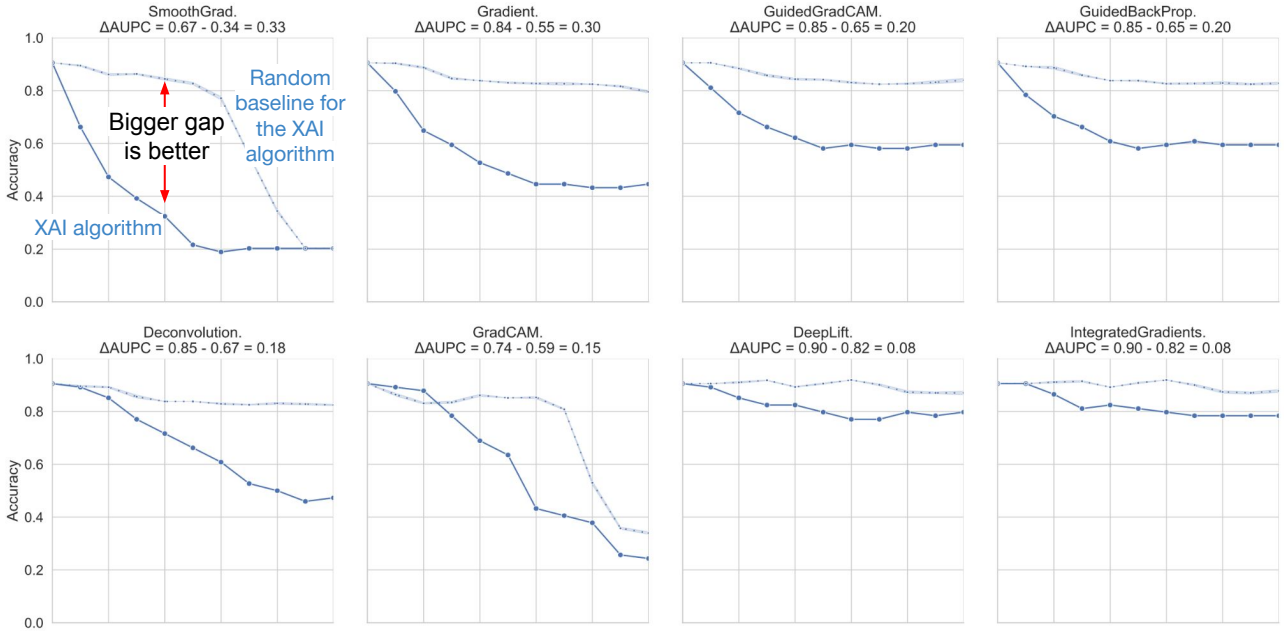


The AI explanation method was selected for being relatively truthful to AI decision process



Assumption:
Removing important features will cause AI performance to drop.

Gradual feature removal experiment



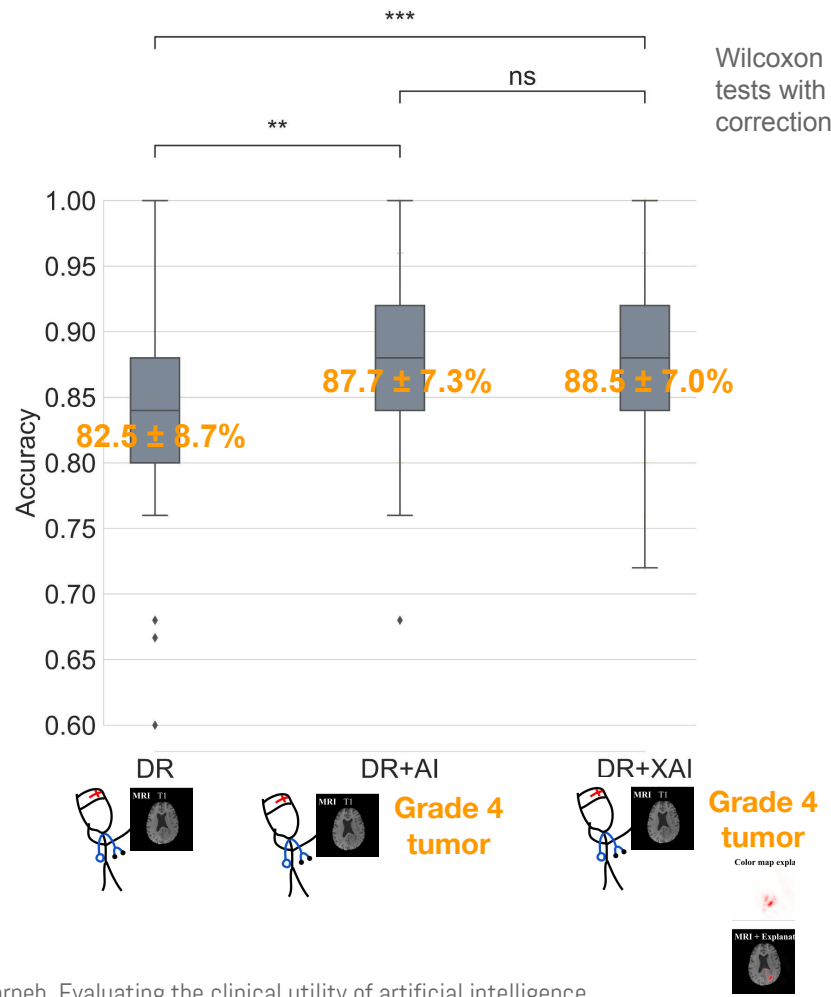
.....

Weina Jin, Xiaoxiao Li, Ghassan Hamarneh. Evaluating Explainable AI on a Multi-Modal Medical Imaging Task: Can Existing Algorithms Fulfil Clinical Requirements? AAAI. 2022

Result

On improving doctors' task performance:

- AI prediction (**DR+AI**) is helpful
- AI explanation (**DR+XAI**) not show additional value



Wilcoxon signed-rank tests with Bonferroni correction

Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

Result

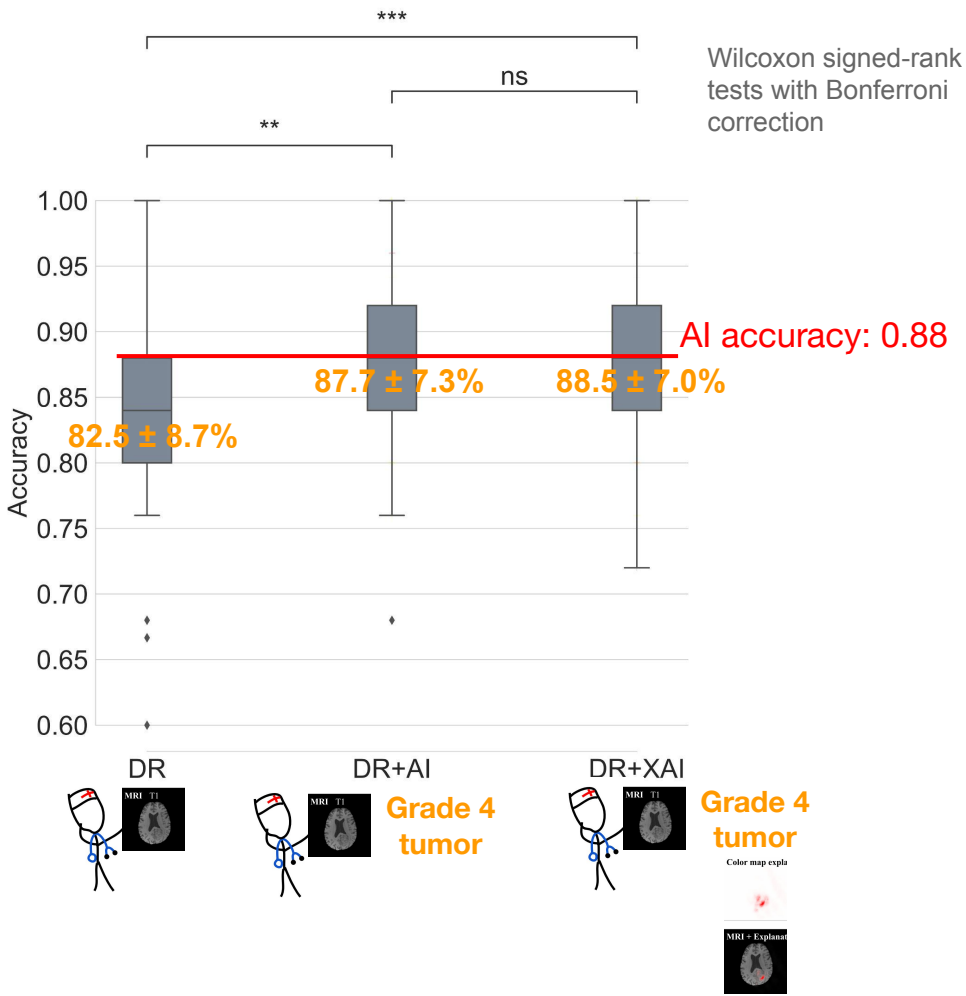
On improving doctors' task performance:

- AI prediction (**DR+AI**) is helpful
- AI explanation (**DR+XAI**) not show additional value

Doctor + AI > max(Doctor, AI) ❌

Suggestion Explanation

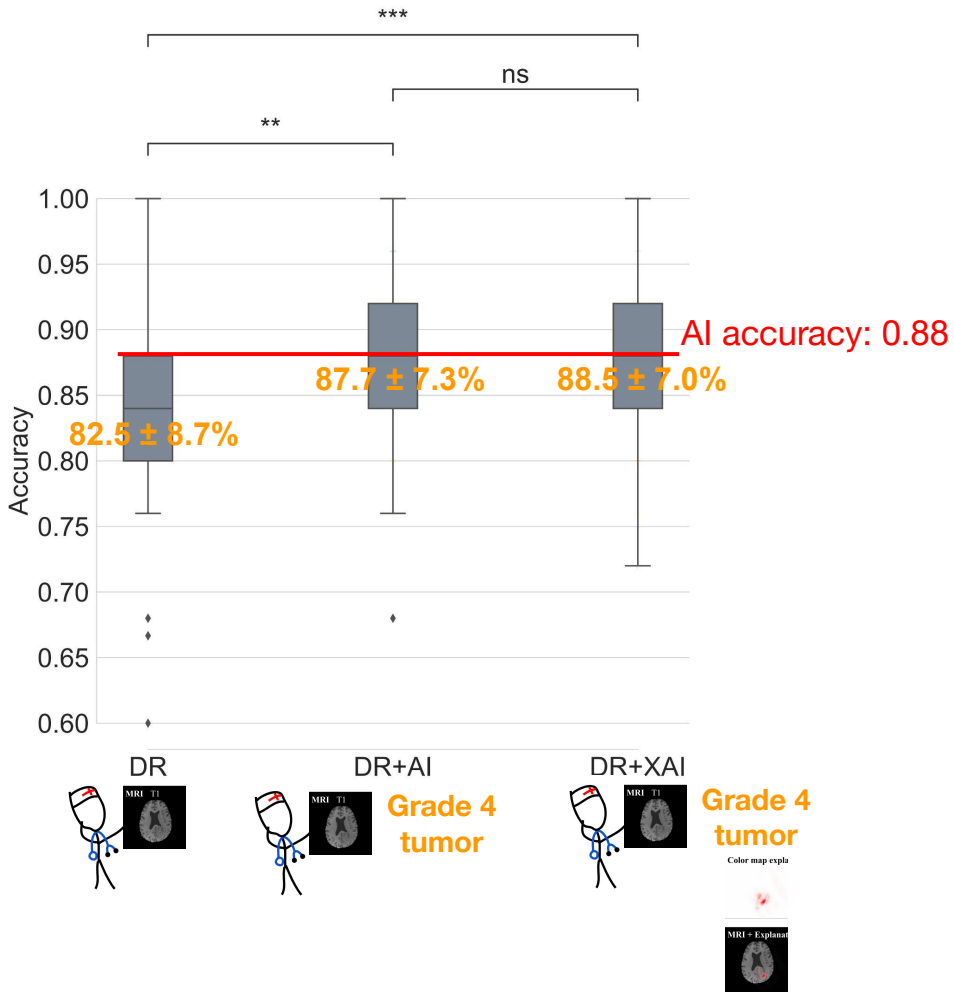
Not achieve complementary doctor-AI performance in **DR+AI** or **DR+XAI**



Critical moment Explainable AI does not work, shall we go back to use black-box AI to replace doctors?

AI to replace humans

Equal work AI



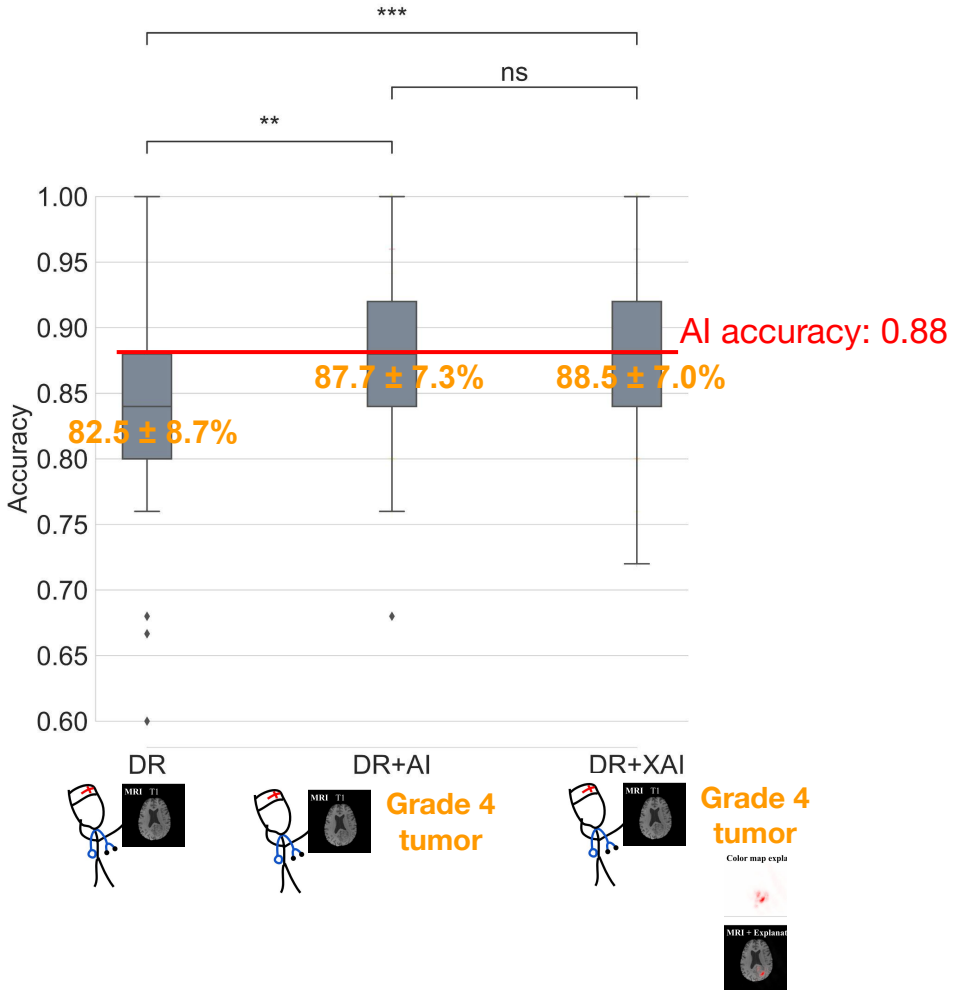
Critical moment Explainable AI does not work, shall we go back to use black-box AI to replace doctors?

AI to replace humans

- Adjust humans to AI
  - Abandon the understanding of human and complexity

Equal work AI

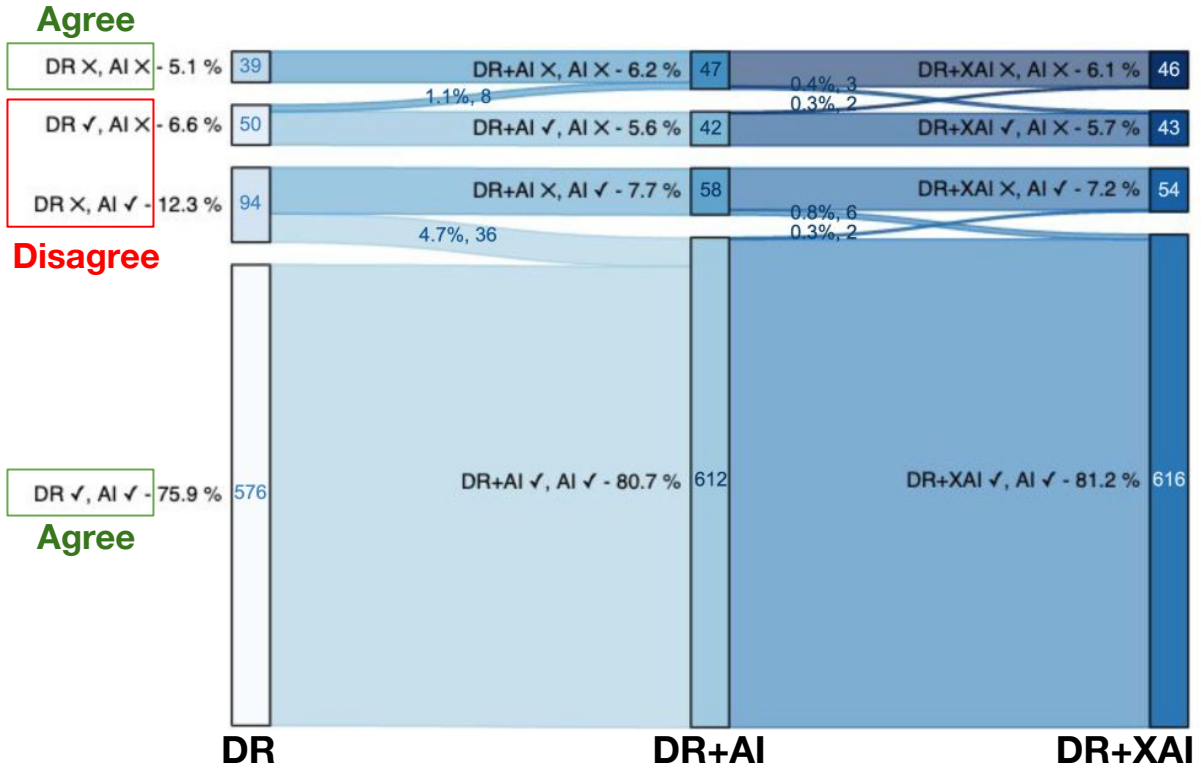
- Adjust AI to humans
  - Embrace uncertainty, diversity, richness, and complexity



Result Why did not AI explanation help?

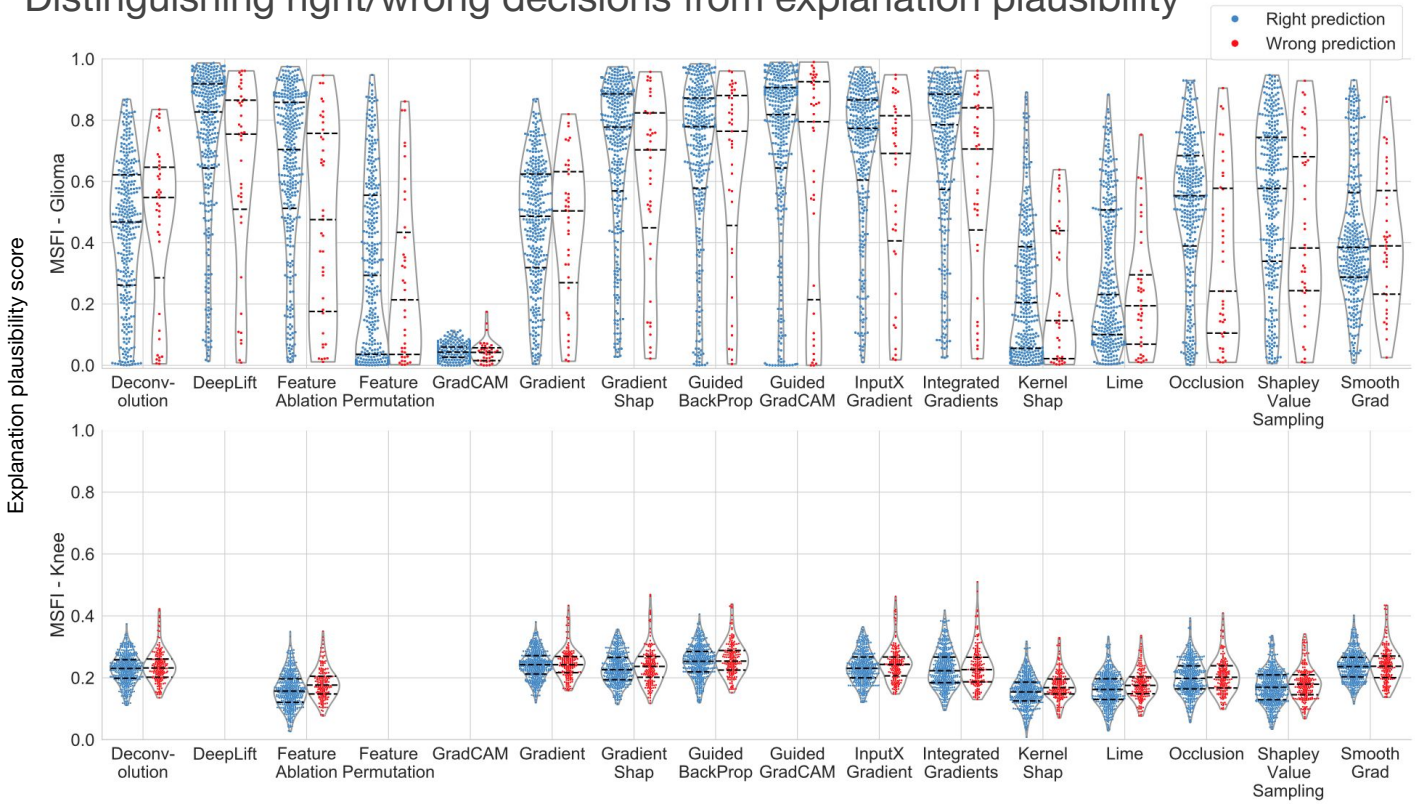
Because physicians had both **changed decisions correctly and incorrectly** with basically **the same amount**.

I.e.: explanation cannot help doctors to discern potentially questionable decisions of AI.



Evaluation of the 16 post-hoc heatmap methods on informative plausibility

Distinguishing right/wrong decisions from explanation plausibility



Human judgment on explanation plausibility can reveal decision quality

Guideline 4

Informative plausibility

W. Jin, X. Li, M. Fatehi, G. Hamarneh, Guidelines and evaluation of clinical explainable AI in medical image analysis, Medical Image Analysis, 2023

What does that (color map region) mean? Like hey, which part of my car gets my car moving? It should say press the accelerator. But yours would just show a dashboard of the car, and show that this button had some red, that button had some red, but it's not an explanation. Let's go to an example, and you'll see, what about the red areas under MRI T1CE (modality)? **Was it central necrosis?** But it couldn't be the central necrosis, because there's more central necrosis in the temporal lobe, and that area didn't get highlighted. So anyway, I don't know, it's just confusing.

…These color maps were totally useless **without text, without any context or explanation**, like those details. The color maps were just pretty, but they didn't explain anything.

– Neurosurgeon #3

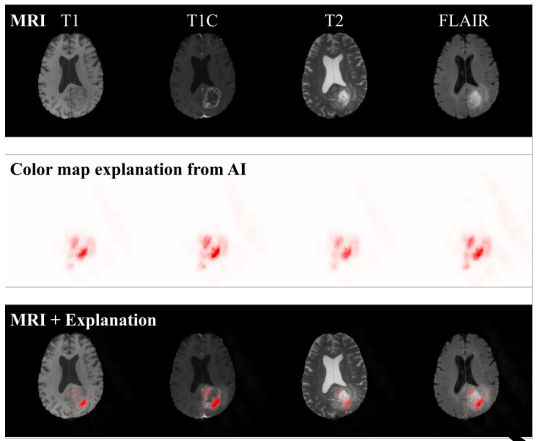
Though the color map is drawing your eyes to many different spots, but I feel like I didn't understand why my eyes were being driven to those spots, like **why were these very specific components important**? And I think that's where all my confusion was.

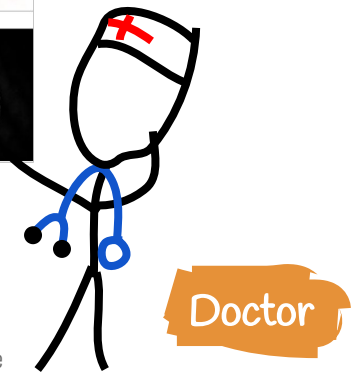
– Neurosurgeon #2

Qualitative results

Why AI explanation did not help?

Because the existing AI explanations do not explain in a clinically relevant way



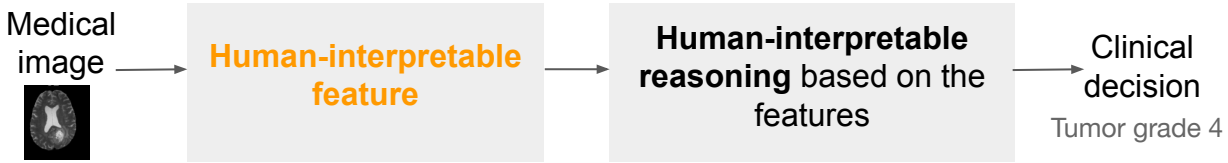


Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

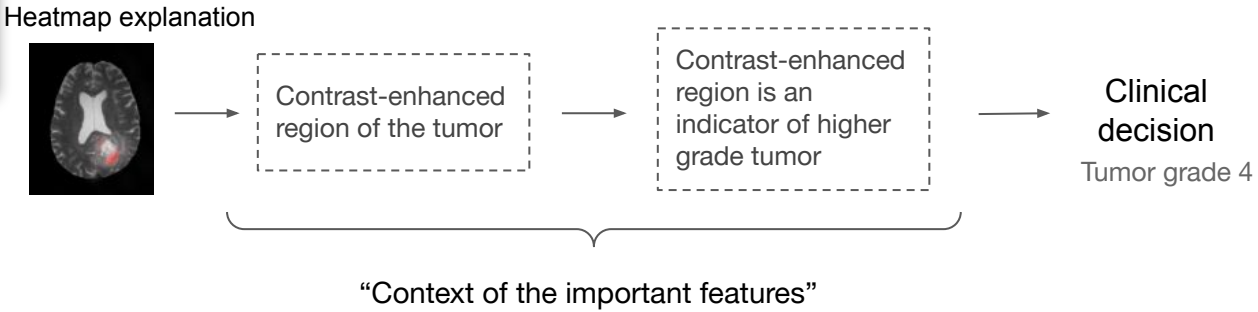
Existing AI explanations do not speak clinical language to explain

> What (explanation) we get currently, when a radiologist read it, they **point out the significant features**, and then they **integrate those knowledge**, and say, to my best guess, this is a glioblastoma. And I have the same expectations of AI (explanation).
>
> – Neurosurgeon #3

Physicians' clinical image interpretation process:

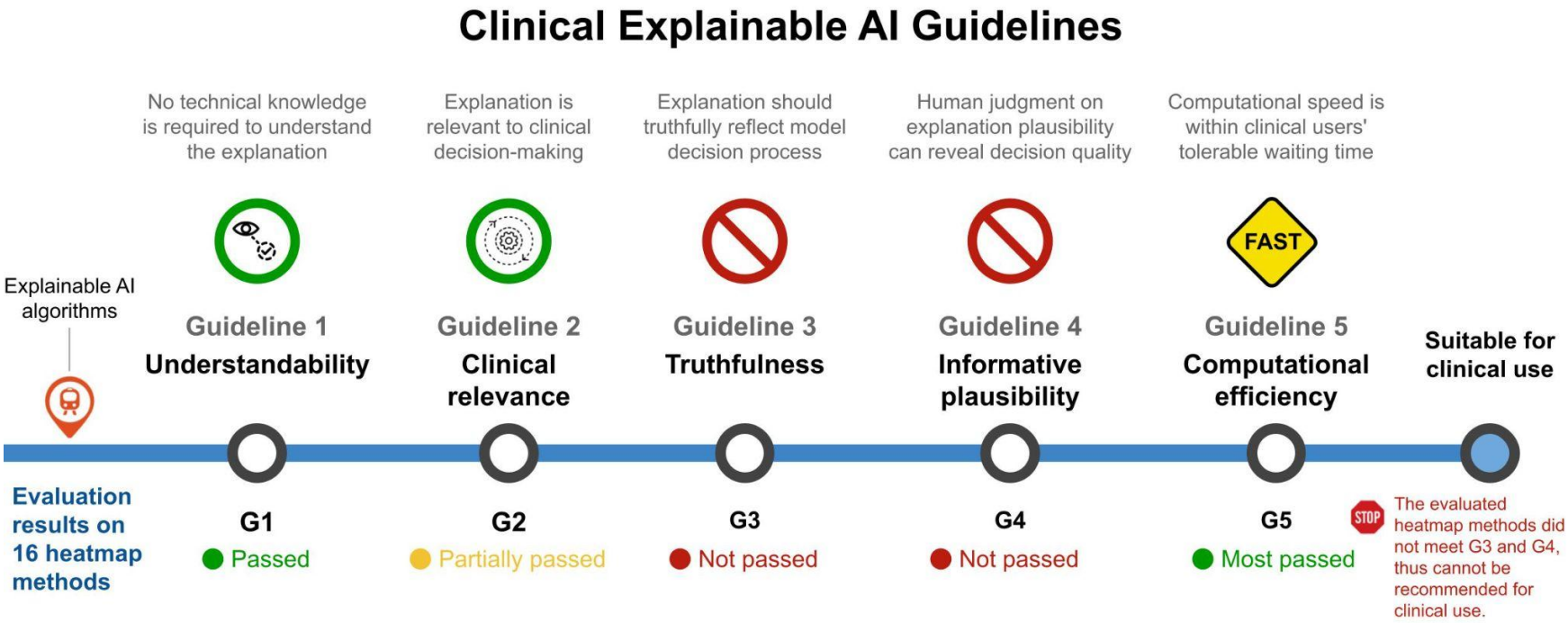


Physicians' interpretation process of AI explanation:



Weina Jin (co-first author), Mostafa Fatehi (co-first author), Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, 2024

The clinical XAI guidelines and evaluation results on 16 heatmap methods



W. Jin, X. Li, M. Fatehi, G. Hamarneh, Guidelines and evaluation of clinical explainable AI in medical image analysis, Medical Image Analysis, 2023

How to achieve complementary human-AI performance?

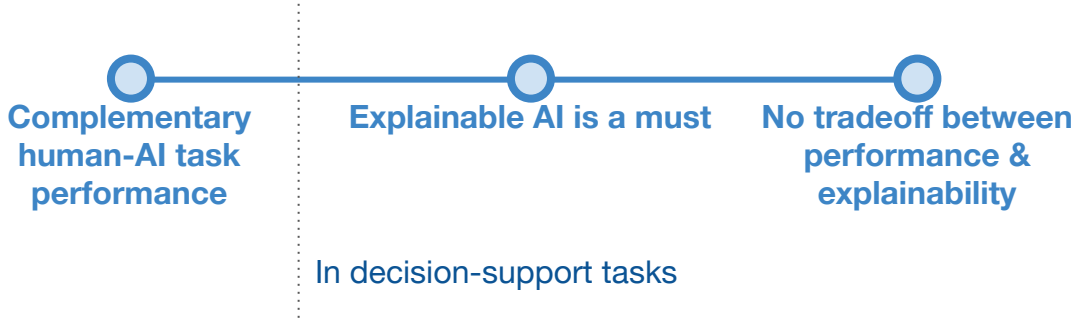
Theorem 2. *Given that the probability of human acceptance of an AI suggestion $f(P_i)$ is a function of the explanation plausibility P_i with a monotonic relationship: $\forall P_i, P_j \in \mathbb{R}$, if $P_i > P_j$, then $f(P_i) > f(P_j)$, if plausibility is indicative of AI decision correctness, with P_i^r and P_i^w denoting the plausibility of AI explanation for the prediction of an instance x_i if x_i is predicted correctly or incorrectly and $P_i^r \neq P_i^w$, then complementary human-AI accuracy can be achieved, $t > \max(h, m)$, conditioned on one of the following:*

1) if $h \geq m$ *and* $\mathbb{E}[f^r] > \frac{h(1-m)}{m(1-h)}\mathbb{E}[f^w]$; *or*

2) if $m > h$ *and* $\mathbb{E}[f^r] > \frac{h(1-m)}{m(1-h)}\mathbb{E}[f^w] + \frac{m-h}{m(1-h)}$

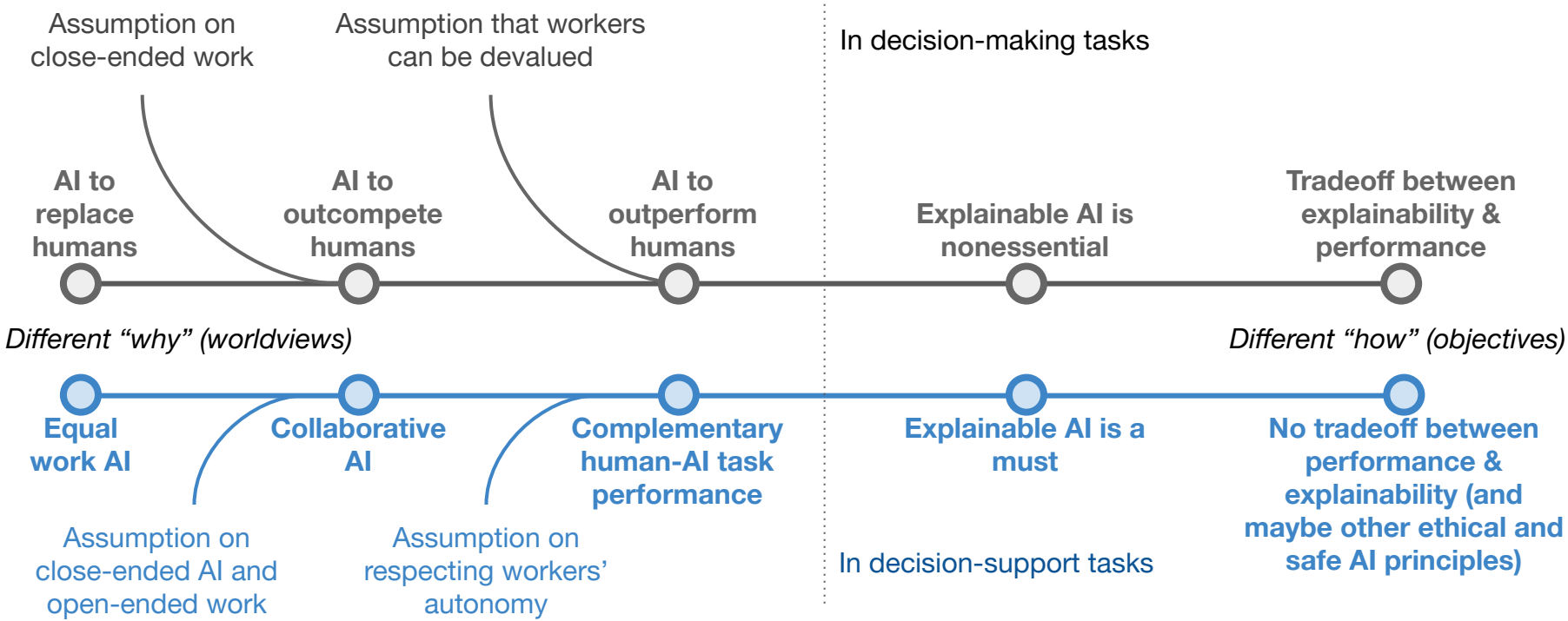
where h, m, and t are the accuracies of the human, AI, and human-AI team; $\mathbb{E}[f^r]$ and $\mathbb{E}[f^r]$ are the expectations of $f(P_i^r)$ and $f(P_i^w)$ over the dataset $x_i \in \mathcal{D}$.

With a **superior or inferior AI**, complementary performance can be achieved, as long as AI can **learn to collaborate with humans** by reliably indicating its decision quality.



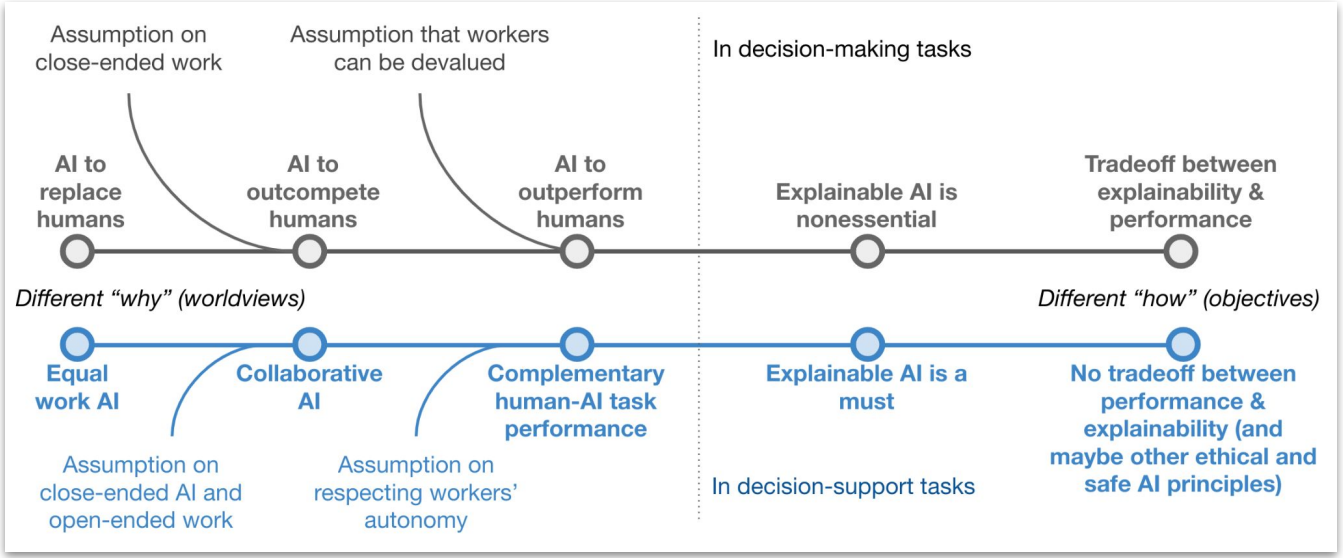
Weina Jin, Xiaoxiao Li, Ghassan Hamarneh. Why is plausibility surprisingly problematic as an XAI criterion? 2024. In submission.

What will AI be under this new “axiomatic system” (worldview)?



Thank you!

Constructing a Different Imagination Beyond "AI Outperforming Humans"



HAMARNEH
Medical Image Analysis Research Group

SFU SIMON FRASER UNIVERSITY

Weina Jin
weinaj@sfu.ca