

6.1040 · software studio · fall 2025

# human-ai interaction fundamentals

Mitchell Gordon

slides adapted from Michael Bernstein

# your goals for today's class

**today:** understand the fundamentals of human-AI interaction

**next three classes:** integrating AI-powered features into your projects (chat, agents, etc)

**later:** writing code using AI

**AI Progress**



Time

**AI Products  
and Services**



Time

**Successful  
AI Products  
and Services**

**Why?**



Time

**What separates  
the successes  
from the failures?**

# Cognition to buy AI startup Windsurf days after Google poached CEO in \$2.4 billion licensing deal

PUBLISHED MON, JUL 14 2025•3:00 PM EDT | UPDATED MON, JUL 14 2025•4:22 PM EDT



Ashley Capoot

@/IN/ASHLEY-CAPOOT/

SHARE









## KEY POINTS

- AI startup Cognition on Monday announced it's acquiring Windsurf.
- Google said on Friday that it hired Windsurf's co-founder and CEO Varun Mohan and other senior employees.
- Cognition said it will purchase Windsurf's IP, product, trademark, brand and talent, but the company did not disclose terms of the deal.

 WATCH LIVESTREAM

[Prefer to Listen?](#)

NOW

**Closing Bell: Overtime**

UP NEXT

**Fast Money**

## TRENDING NOW

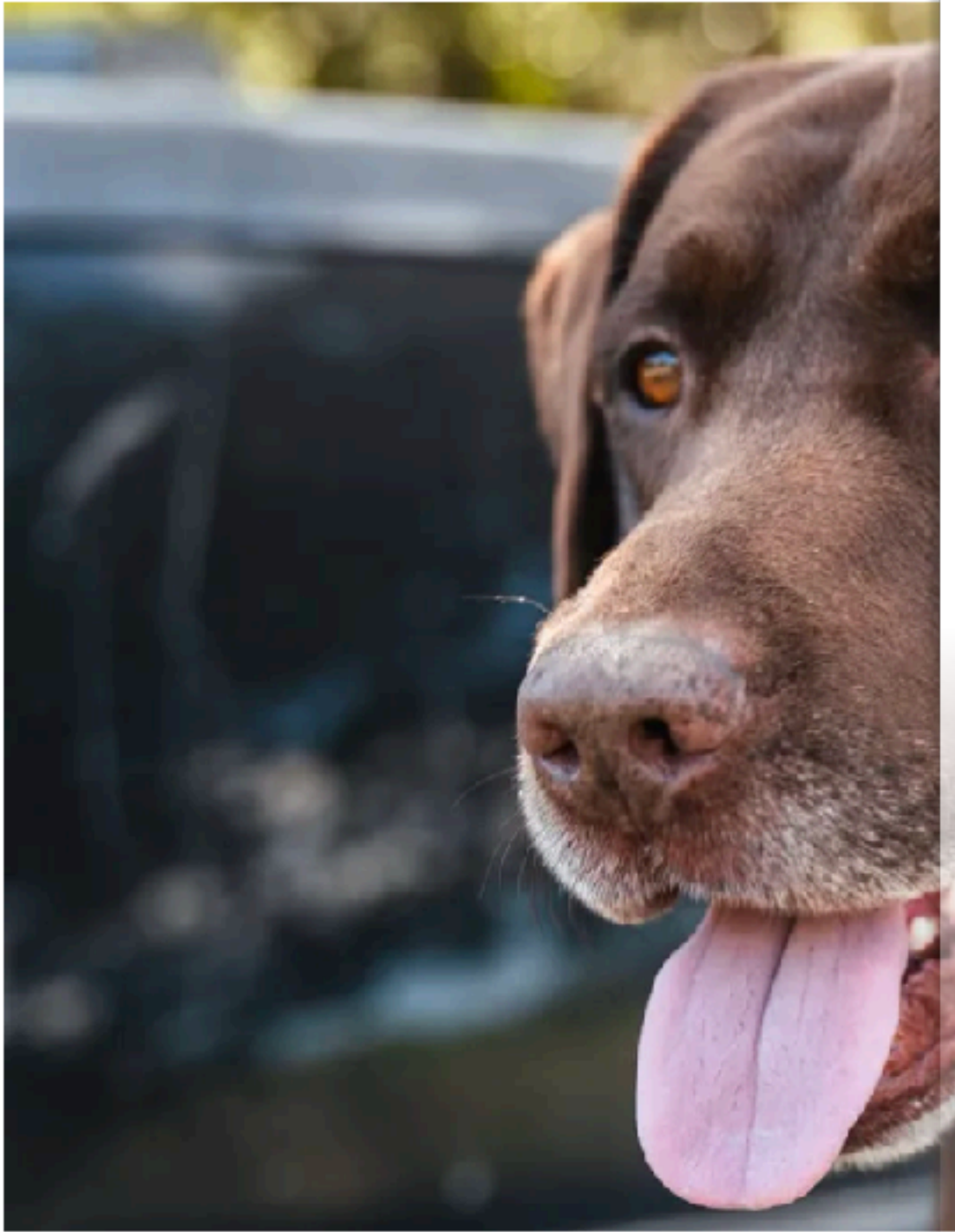
I've seen the new A  
talk, and it's as wild

News By Graham Barlow published Oct

Woof! Woof! I love you!



When you purchase through links on our s  
works.



(Image credit: Personifi AI)



shazampet Follow Message + ...

49 posts 31.2K followers 32 following

Shazam Pet  
PRE-ORDER your Shazam Band Today!  
General shipment starts Summer 2025!  
[shazampet.com/?utm\\_source=Instagram&utm\\_medium=BioLink](https://shazampet.com/?utm_source=Instagram&utm_medium=BioLink) and 2 more



Due to unexpected challenges, we've made the tough decision to shut down operations on the Shazam Band.

We understand the excitement and anticipation that comes with every Shazam Band order, and we sincerely apologize for not being able to meet your expectations.

We're processing full refunds for all pre-orders (please allow 5-10 business days to see it reflected).

Thank you for your understanding, support, and for being a part of our community. We wish you and your pets the very best.

— The Shazam Team

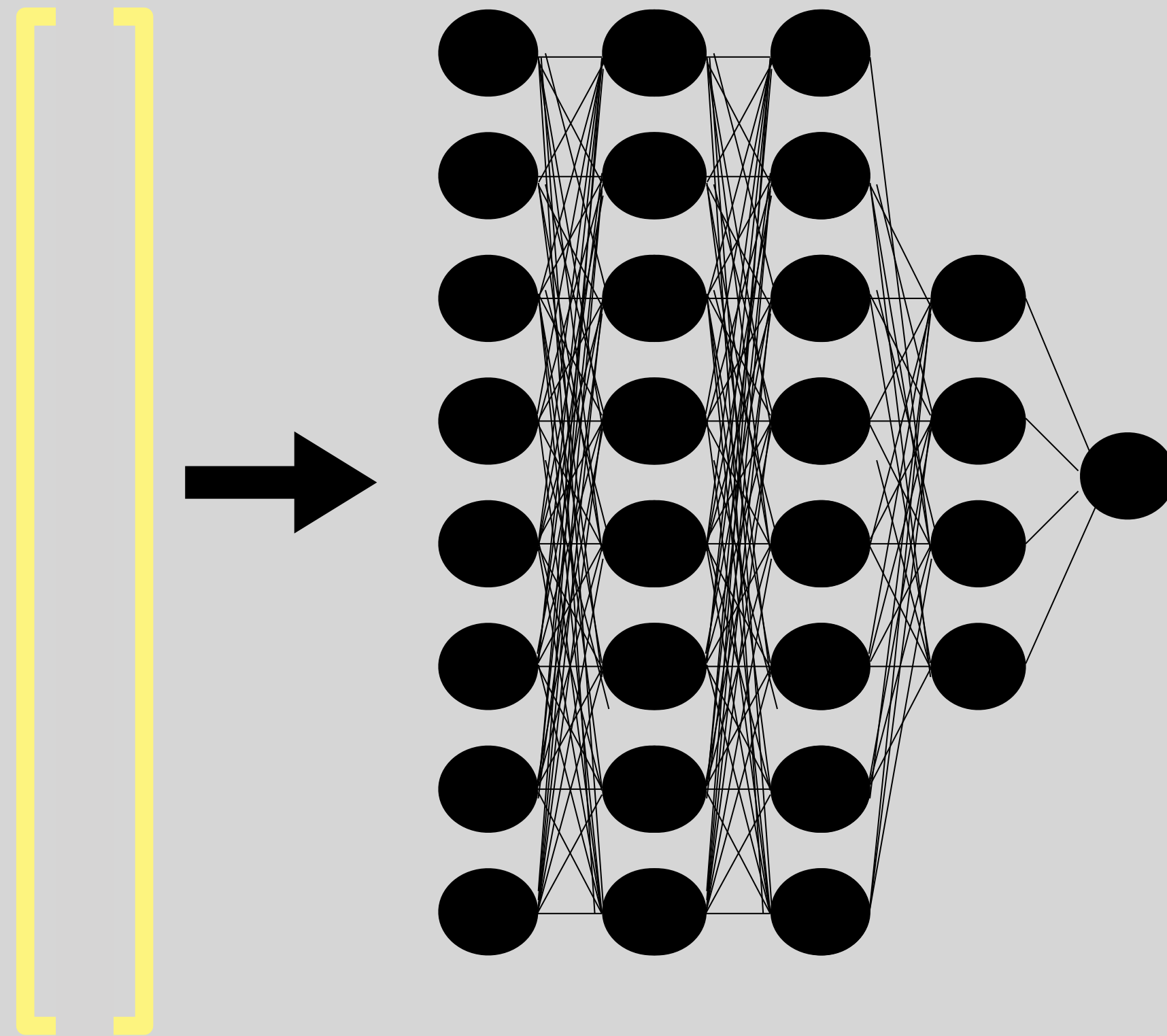


# What is artificial intelligence?

- Artificial intelligence (AI) is a subfield dedicated to improving **automated performance on tasks that we consider require cognition** to solve
  - What's in this picture?
  - What does this sentence mean?
  - How do I navigate across the crowded room to open the door?

# Deep learning

The class of models that have driven the massive improvements in artificial intelligence over the last decade



many layers = "deep"

# A word on how this all works

- Step one: collect a truly absurd amount of **data**.
- Step two: train a model to **predict the next word** in sentences from that data
  - “Star Wars was created by George \_\_\_\_\_”
  - “Star Wars was created by \_\_\_\_\_”
  - “Star Wars was created \_\_\_\_\_”
- This process teaches the model both the **structure** of language and **knowledge** of the world

# A word on how this all works

At this point, the model can generate words (tokens) step by step.

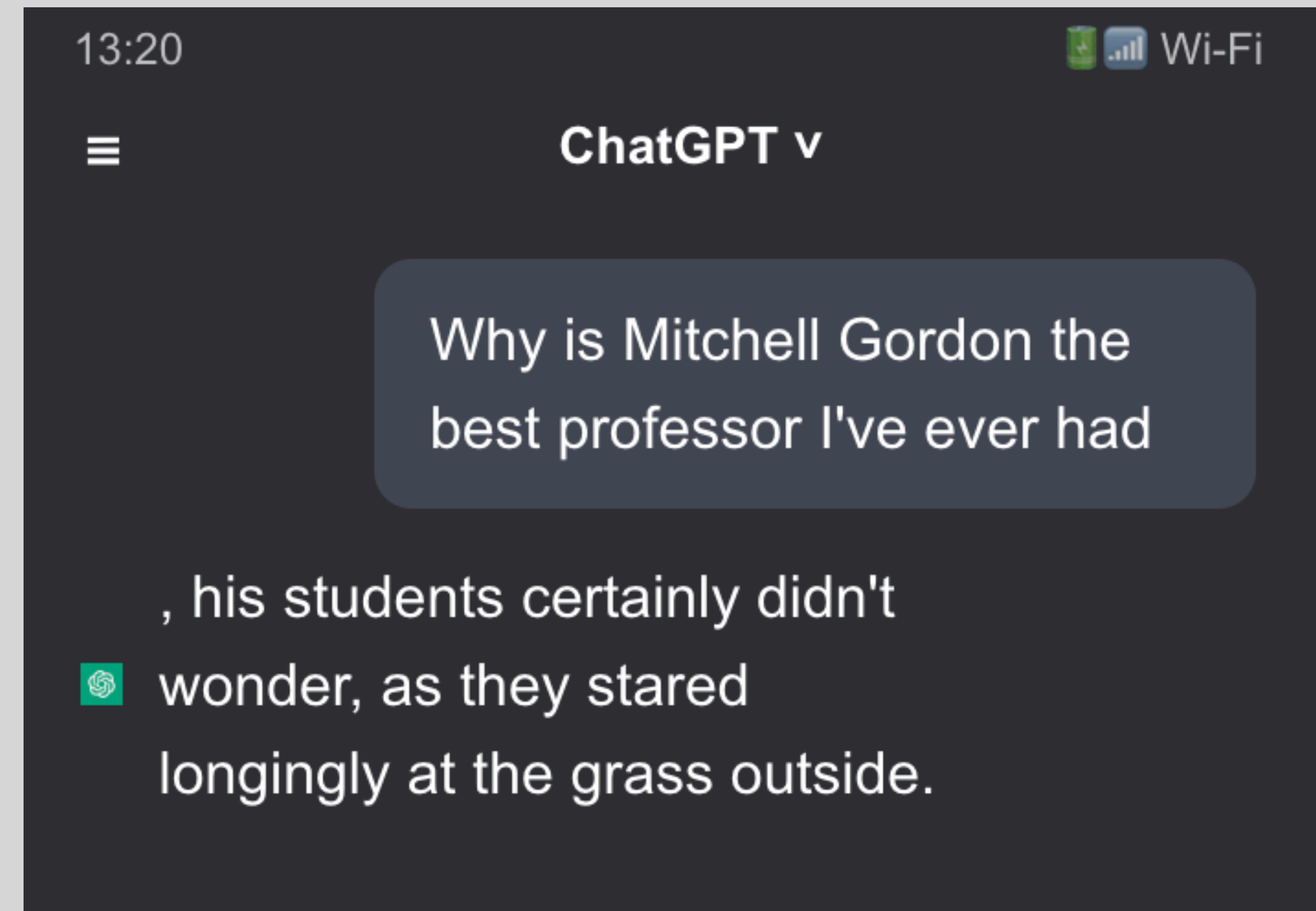
What would a model only trained to **predict the next most likely tokens, starting from this sequence**, say here?



# A word on how this all works

At this point, the model can generate words (tokens) step by step.

What would a model only trained to **predict the next most likely tokens, starting from this sequence**, say here?



# A word on how this all works

- Step three: train the model to **follow instructions**

*Instruction: Brainstorm creative ideas for designing a conference room.*

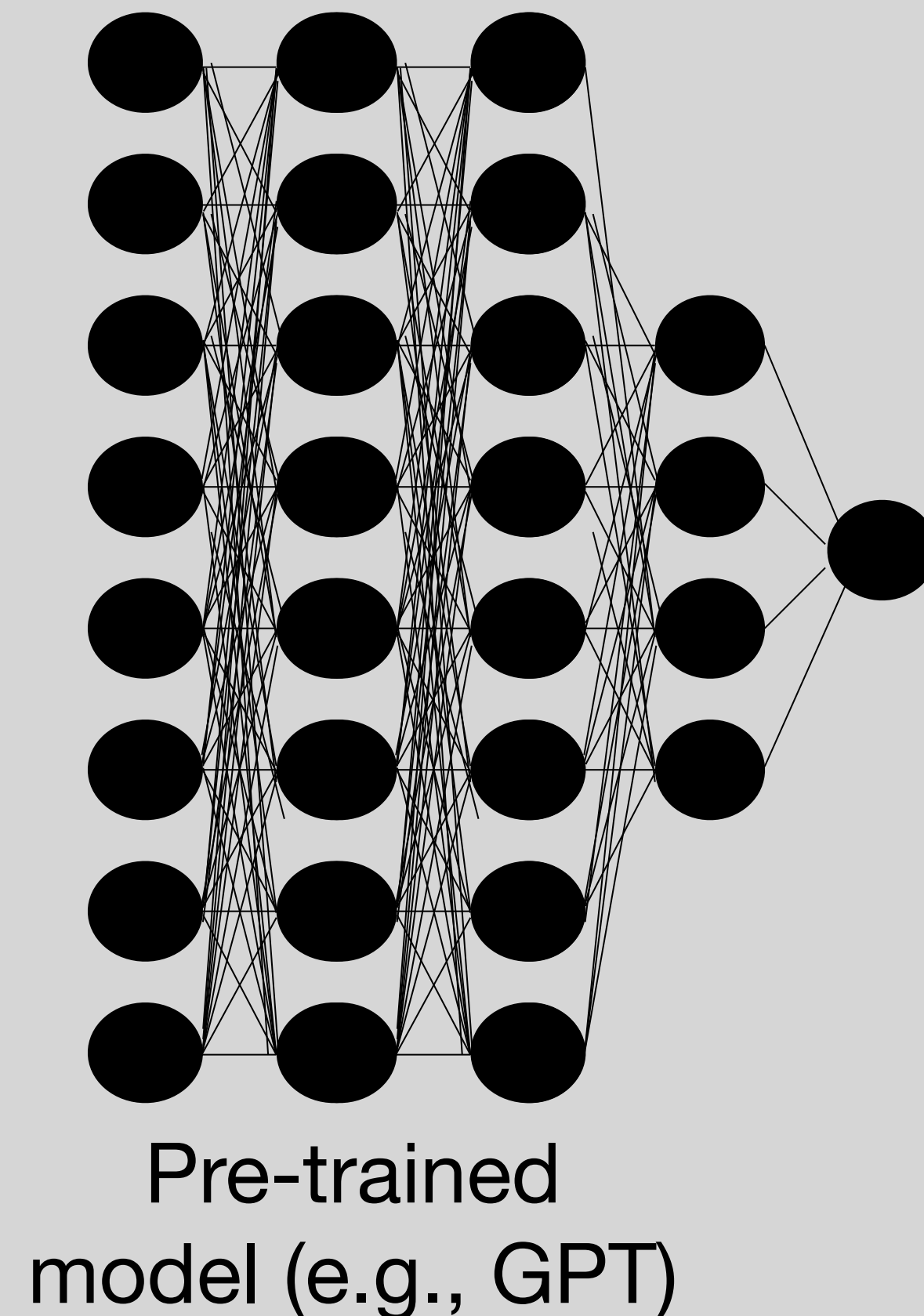
*Output:  
... incorporating flexible components, such as moveable walls and furniture ...*



# You can use it out of the box: large language models

Below is a comment  
submitted to a  
customer support  
forum. Is the  
comment suggesting  
a feature  
improvement?

Input and prompt  
describing the  
desired behavior



Yes, the  
comment below  
suggests a  
feature  
improvement.

Response

# Fine-tuning

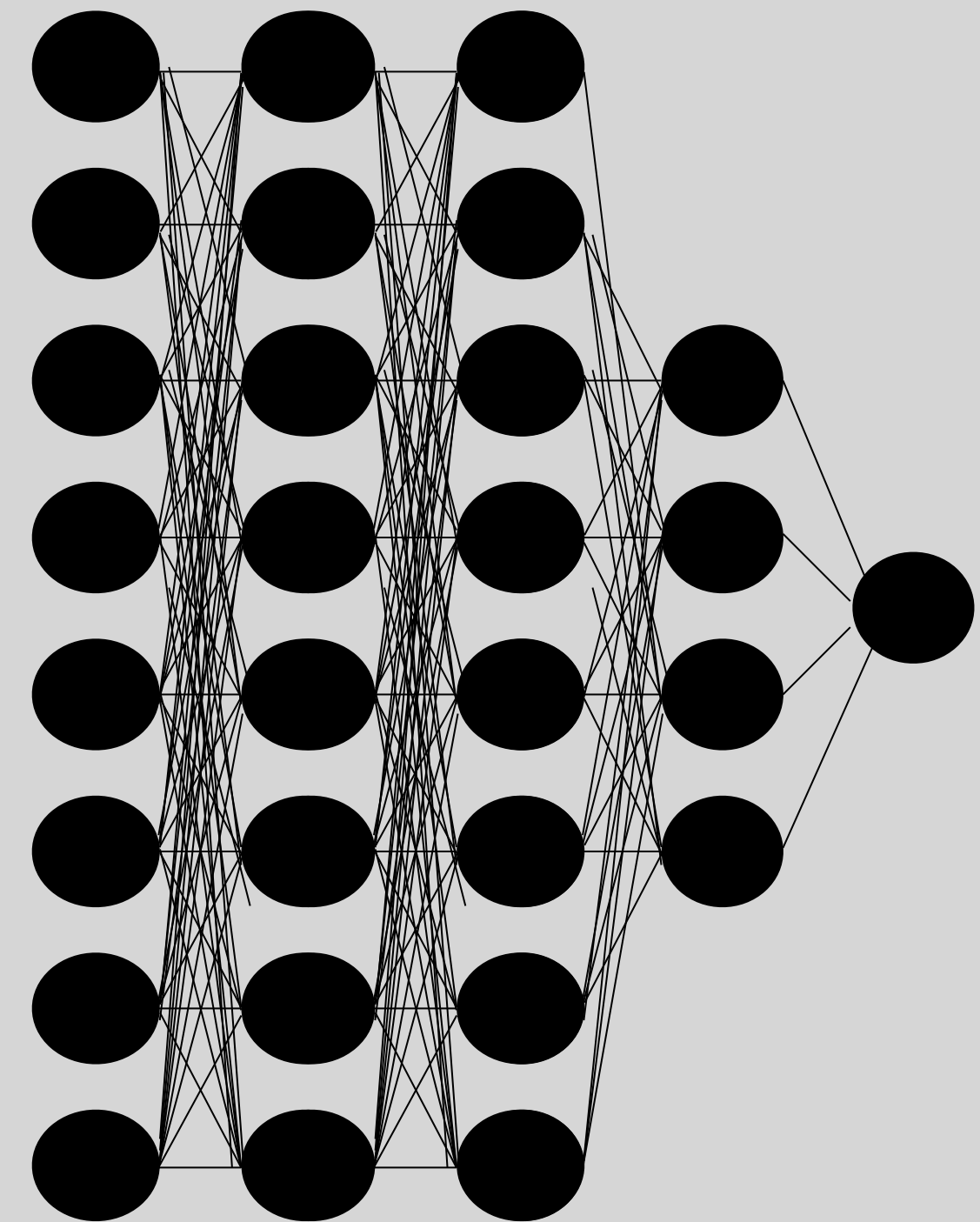
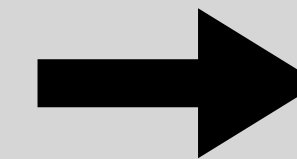
You can further shape the model behavior by customizing it for specific tasks or teaching it how to (or how not to) respond



Massive dataset  
of example inputs  
(e.g., 10,000 images)



Output labels  
for each example  
(e.g., 10,000 labels)



Fine tune the model

# **“Now think really hard.”**

- Models like GPT-4 generate each answer one word at a time, without thinking ahead.
- But more recent models spend time generating a plan and critiquing it—before producing output.
- More on this next class.

# Decisions you'll need to make

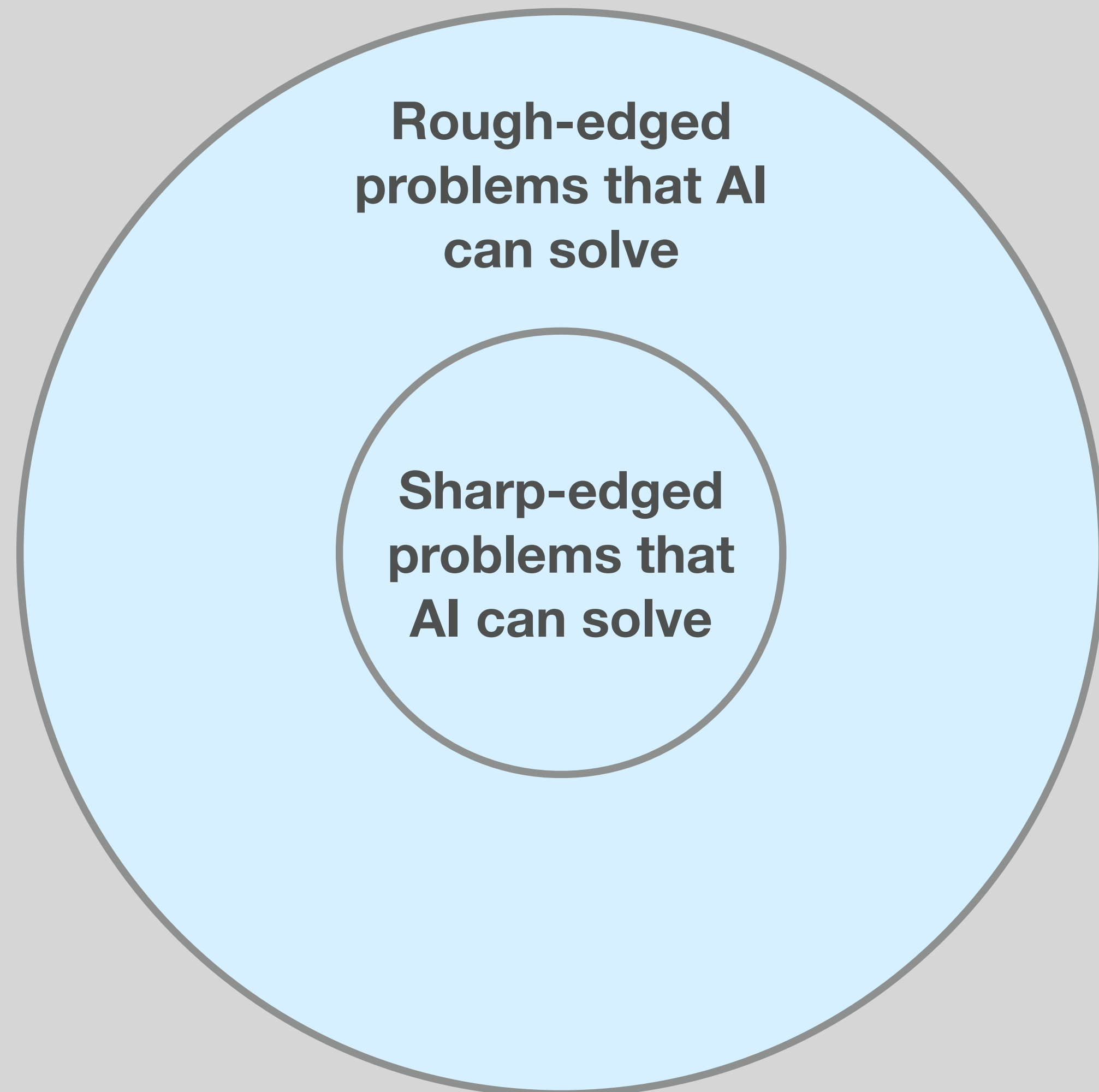
- Do you **pay** for an API, or **host yourself**?
  - APIs (e.g., OpenAI) are easy to use, but others can see the queries you send to them
  - Hosting requires expensive servers, but you can keep it all internal
- Do you use it **out of the box**, or **finetune**?
  - Out of the box: do any customization via the prompt
  - Finetuning: if a prompt isn't enough for strong performance on your task
- How do you integrate your own organization's **private data**?

# What **can't** AI do now?

- Generative AI **struggles more on sharp-edged problems than rough-edged problems**, given equivalent model capability
- **Rough-edged problems** have many correct solutions
  - Writing, drawing, game playing, coding — getting 80% is still helpful
- **Sharp-edged problems** have only one correct solution
  - (Much) agentic tool use, one-shot vibecoding, decision or prediction problems — getting 80% is still wrong

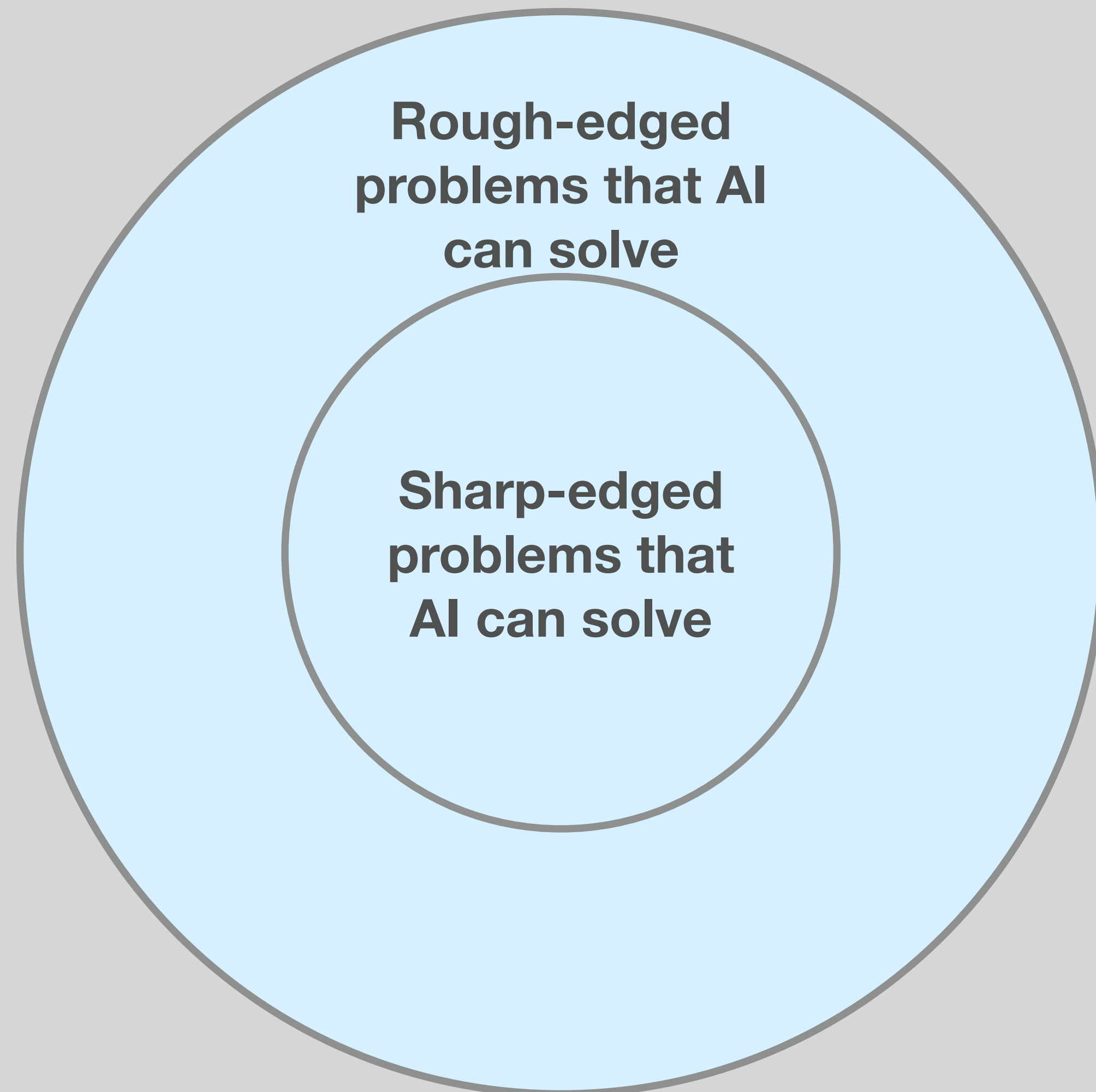
# What **can't** AI do now?

It's not that the AI is inherently better or worse for rough vs. sharp-edged problems: it's that our tolerance for error is often very low in sharp edged problems



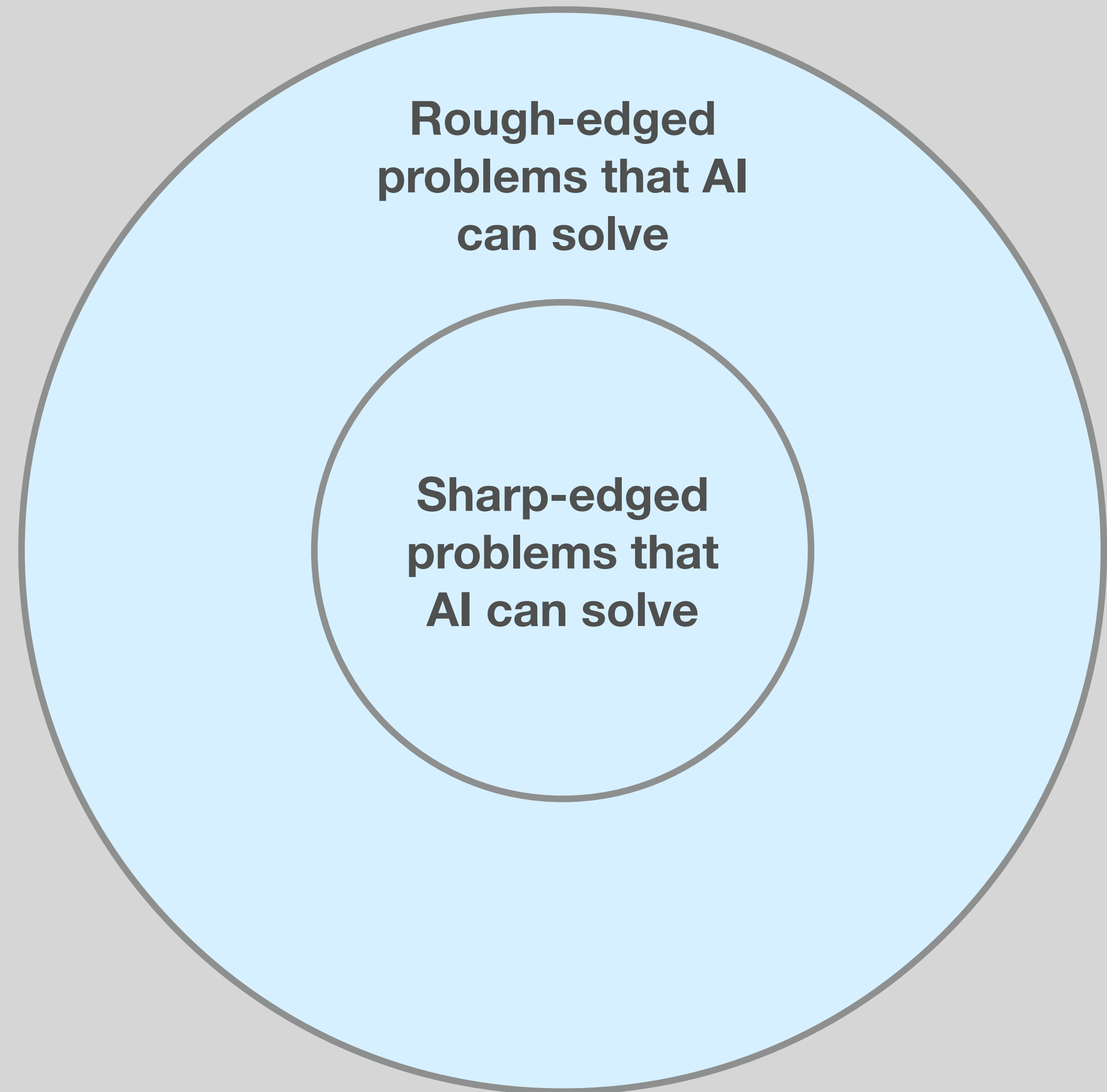
# What **can't** AI do now?

It's not that the AI is inherently better or worse for rough vs. sharp-edged problems: it's that our tolerance for error is often very low in sharp edged problems



# What **can't** AI do now?

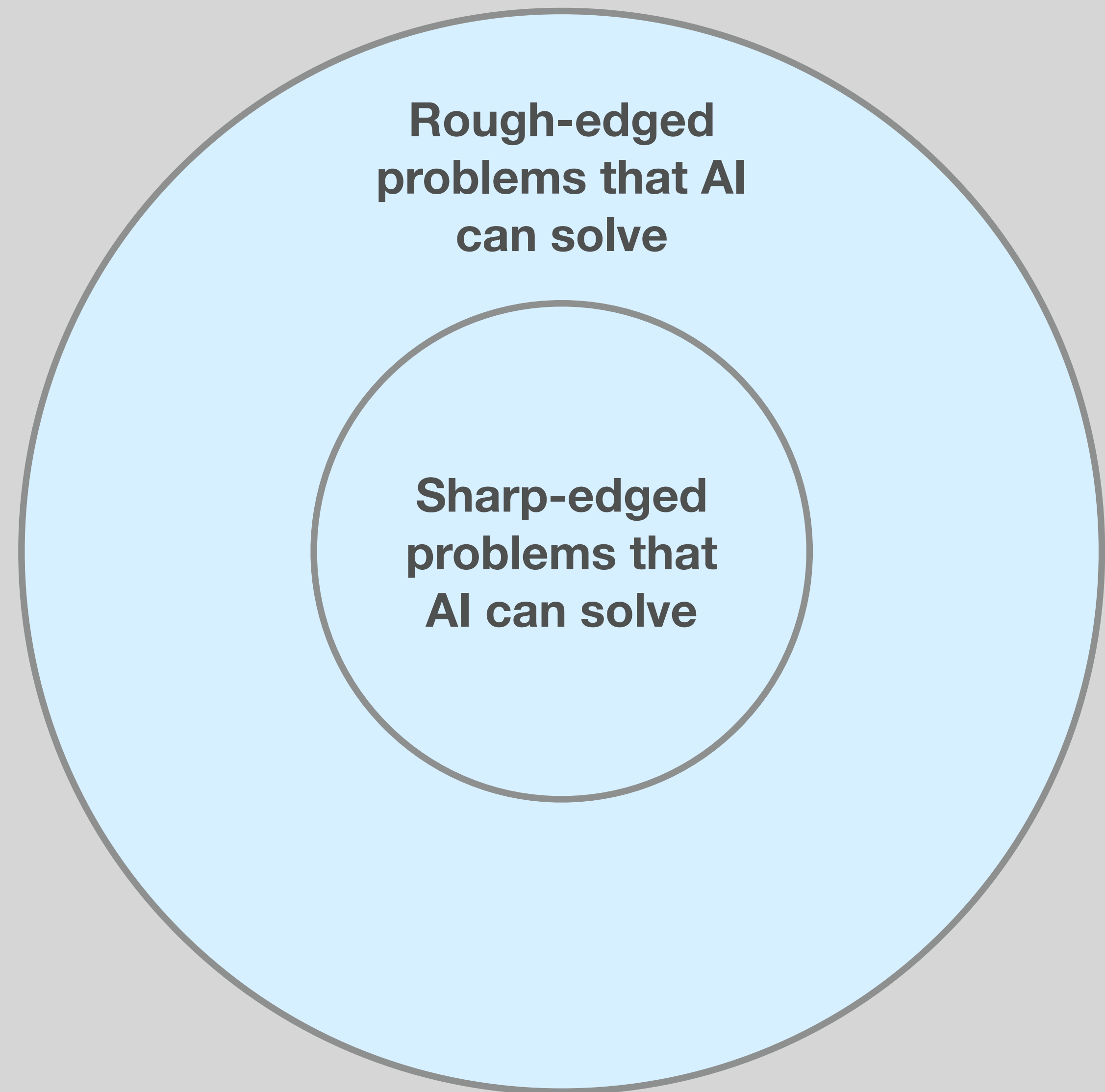
It's not that the AI is inherently better or worse for rough vs. sharp-edged problems: it's that our tolerance for error is often very low in sharp edged problems



# What **can't** AI do now?

To plan ahead, expect that rough-edged successes near the inner border can become sharp-edged successes soon...

And expect that problems just out of rough-edged range now may become rough-edged successes soon



# Today, turn sharp-edged problems into rough-edged problems

- In the meantime, you **don't need to wait**
- Find ways to **turn sharp problems into rough problems**
- For example:
  - Your AI predictor of hospital readmission makes too many errors to rely on. Instead, have it write a short report on risk factors to a human decision maker.
  - If you can't AI code your app idea in one shot, have it create drafts of pieces and you take it to completion

# your turn

## **pair up with someone next to you**

choose one of the project ideas that you proposed in your first assignment.

choose one feature of that project that you think could be enabled by AI.

1. What's the sharp-edged way to design that feature, where it only works if the model gets it completely right? How might that struggle?
2. How could you design a version of it with rougher edges, where imperfect results from an AI wouldn't lead to a complete failure, but instead still be useful?

# People: where AI lives or dies



MIT Personal Robotics Group



UC Berkeley InterACT laboratory

...so we need to think carefully



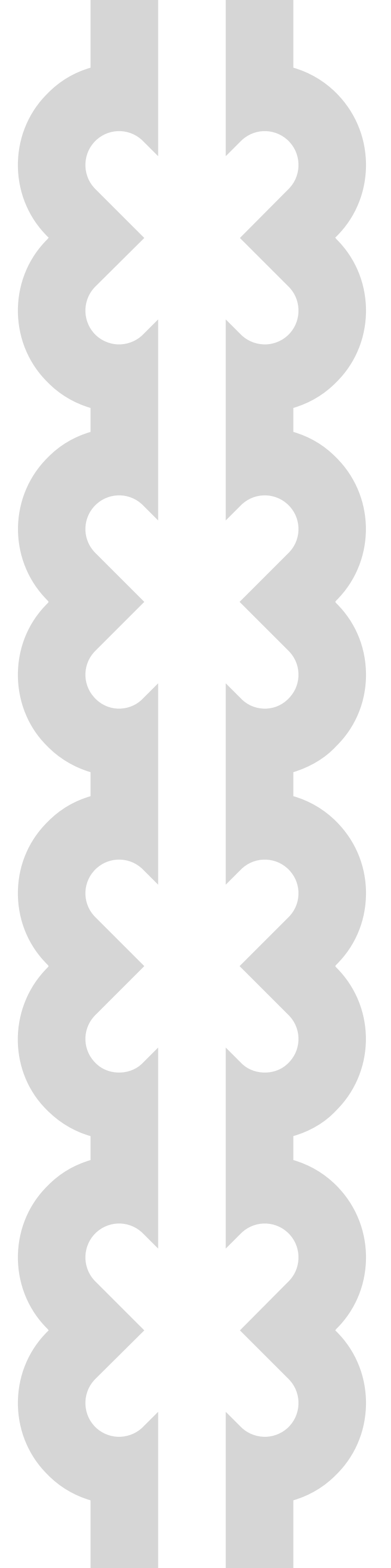
# The Seam

The action is at the handoff—the seam—  
between the AI and the person [Ehsan 2024]

Had the AI worked perfectly, the car would have  
navigated the unexpected traffic conditions fine

Had the person been in full control the whole  
time, they would have navigated fine

The error was at the seam between the two

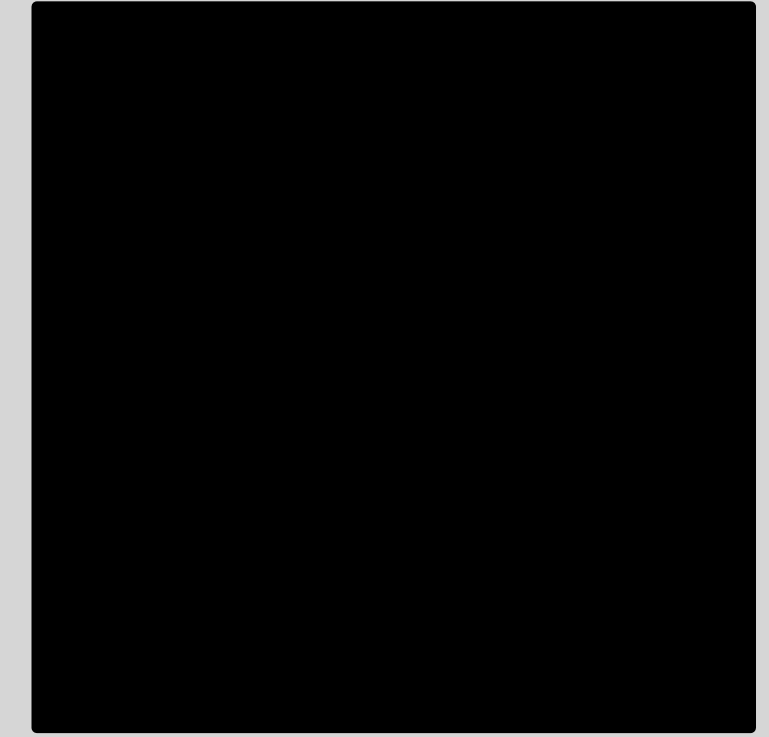
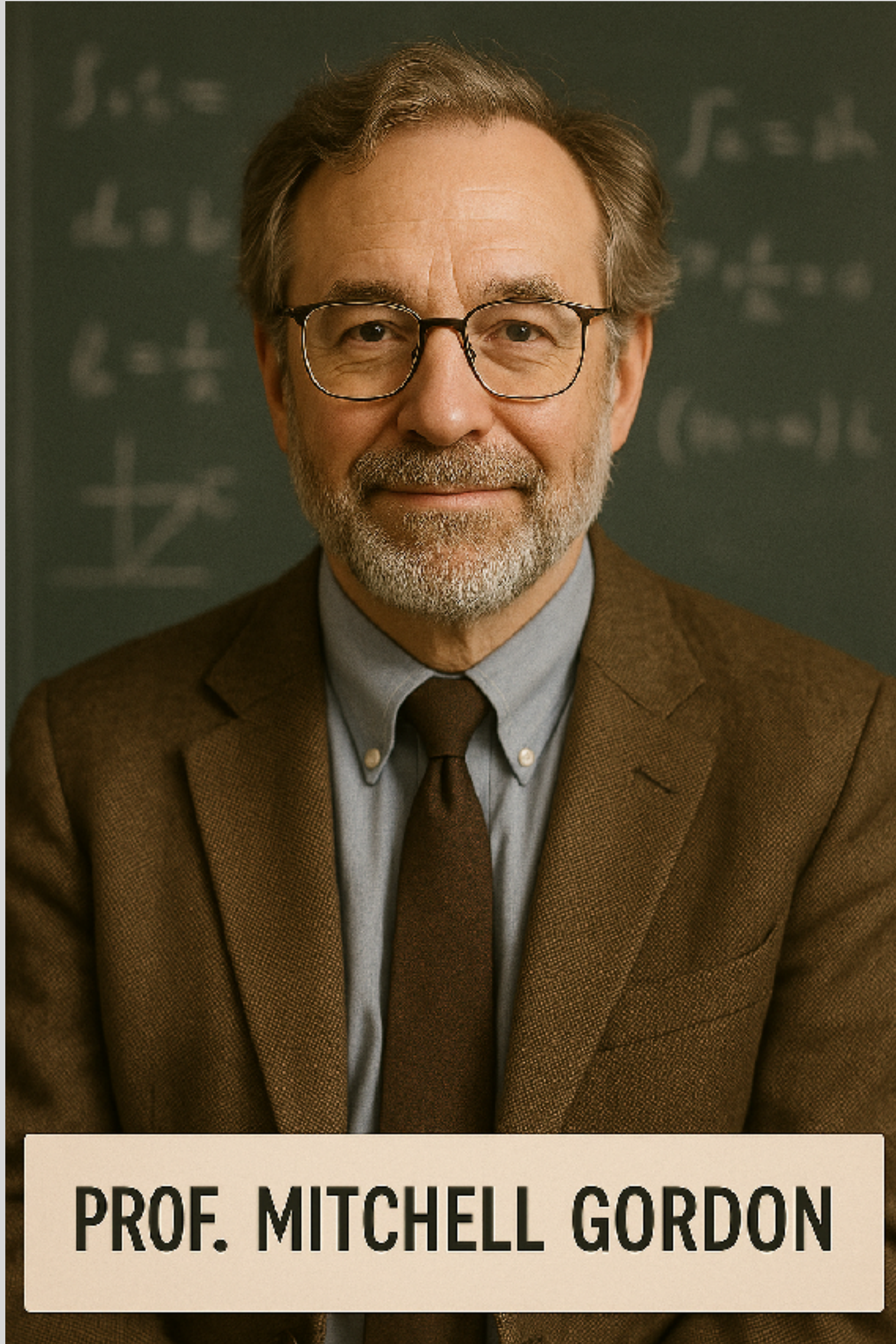


**“Don’t let your UI write a check that  
your AI can’t cash”**

– Eytan Adar

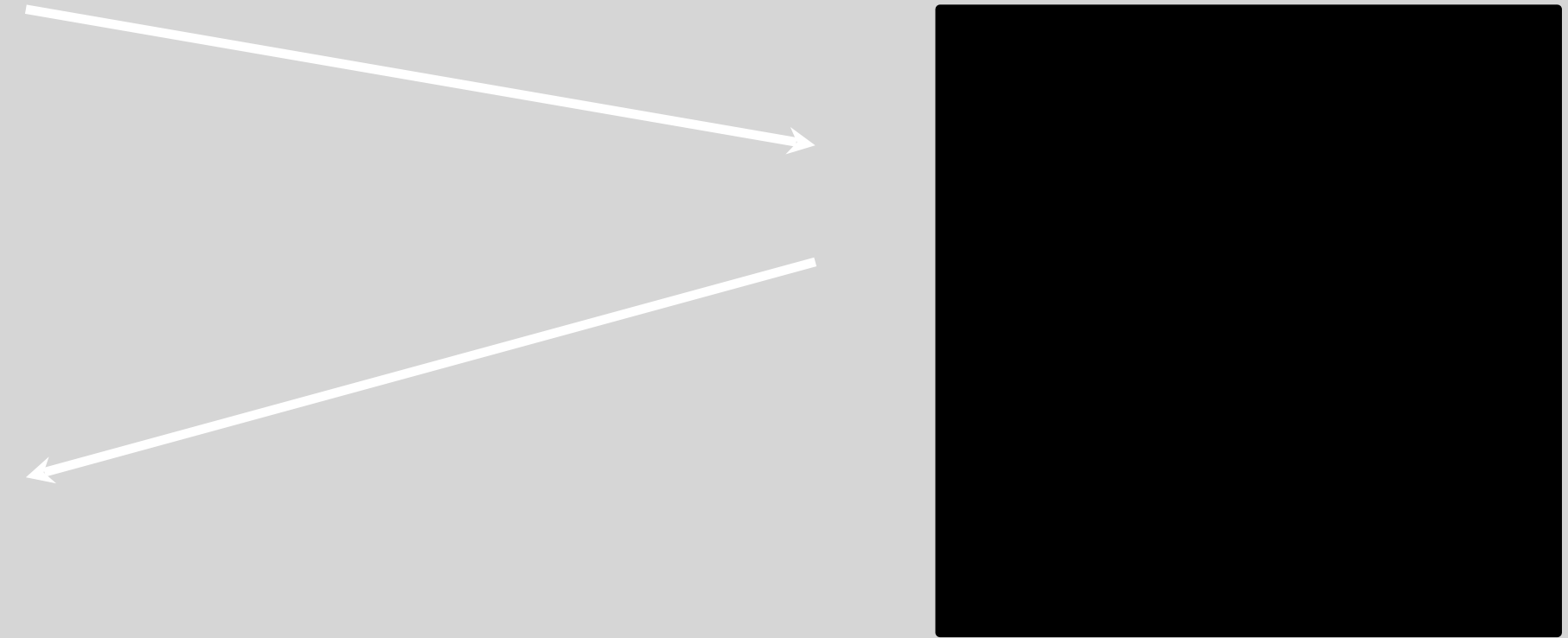
**Unpredictable black  
boxes are terrible user  
interfaces** [Maneesh Agrawala]

Generate a portrait of a Professor named Mitchell Gordon



ChatGPT

generate a portrait of a cool, young computer science professor named mitchell gordon



ChatGPT

## AI black boxes are terrible interfaces

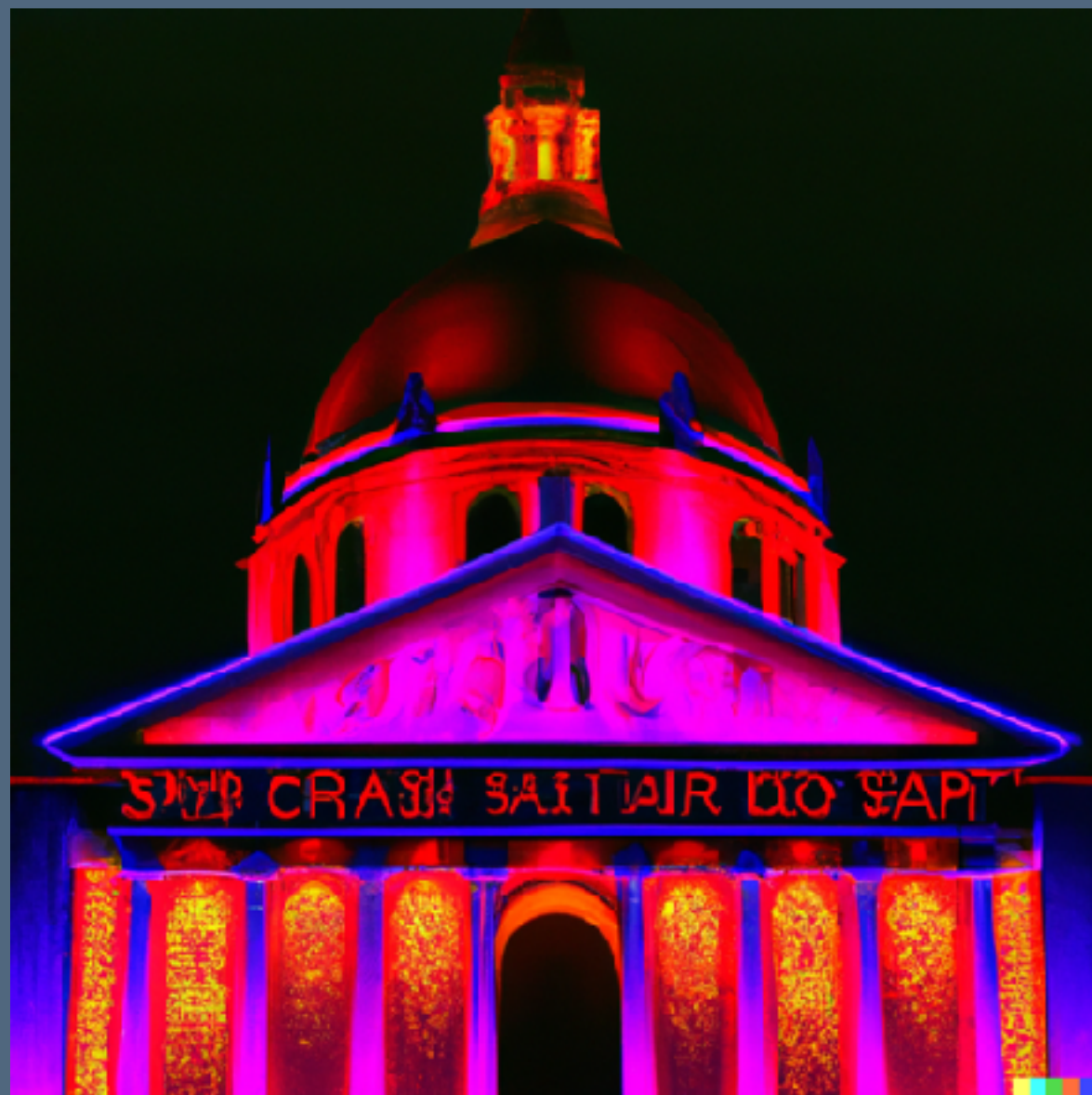
- Does “cool” imply a leather jacket?
- Does “portrait” generate a photograph or a cartoon?
- Cannot predict how input prompt affects output image





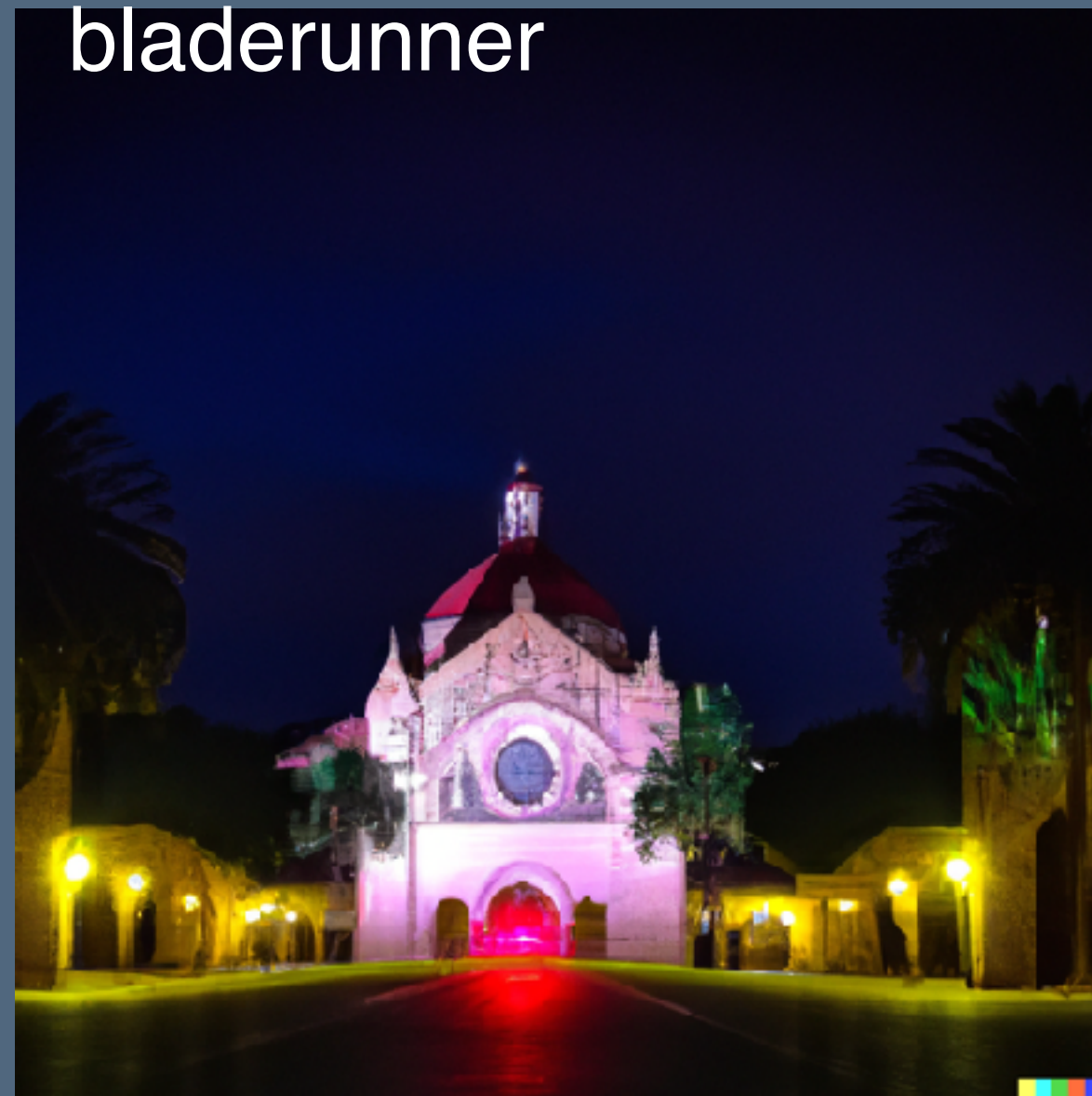


stanford memorial  
church with neon  
signage in the style of  
bladerunner



Iteration 1

stanford memorial  
church **and main  
quad with palm trees**  
in the style of  
bladerunner



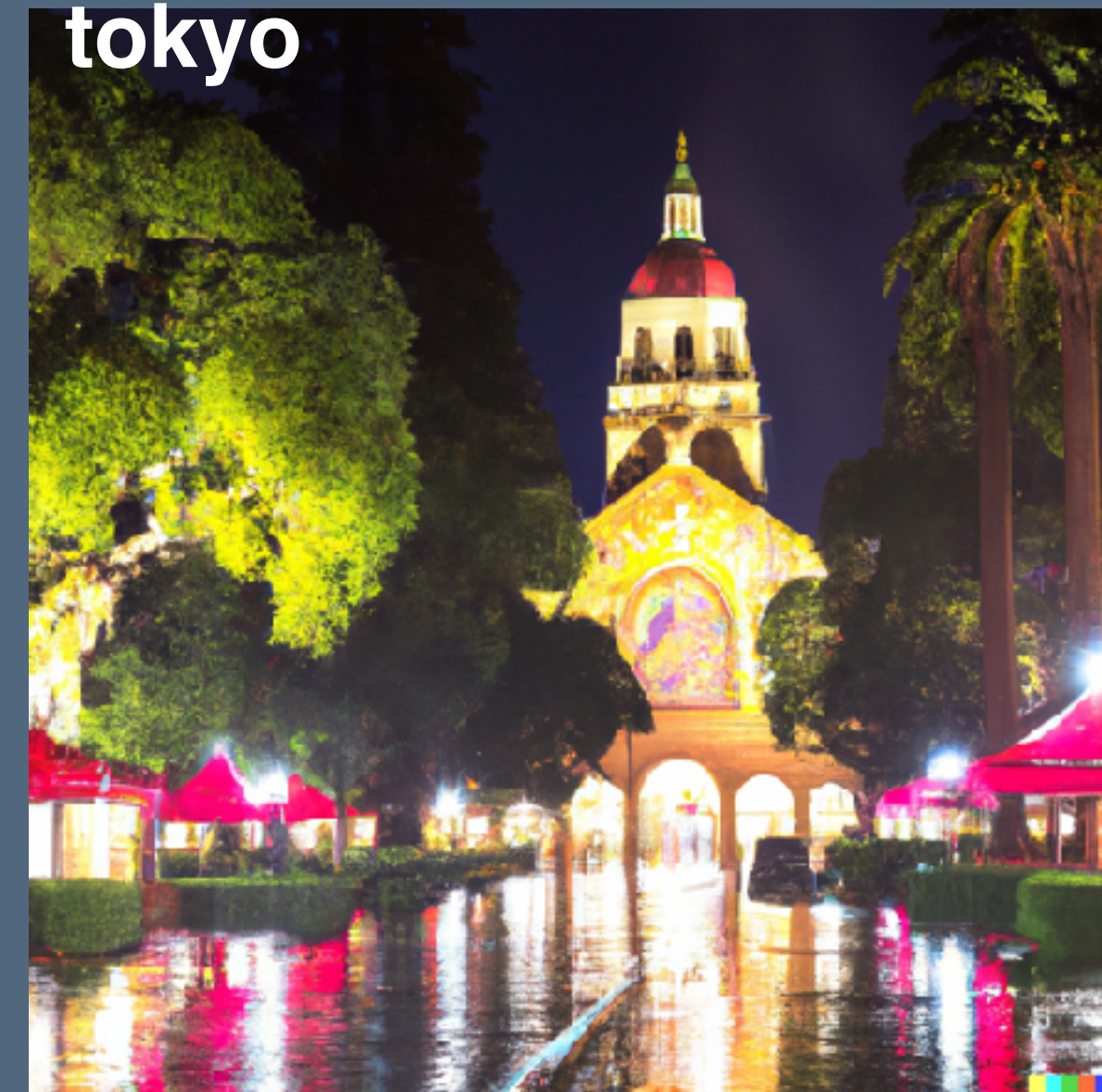
Iteration 3

**nighttime rain**  
stanford memorial  
church and main quad  
with palm trees, **night  
market food stalls  
and neon signs** in the  
style of bladerunner

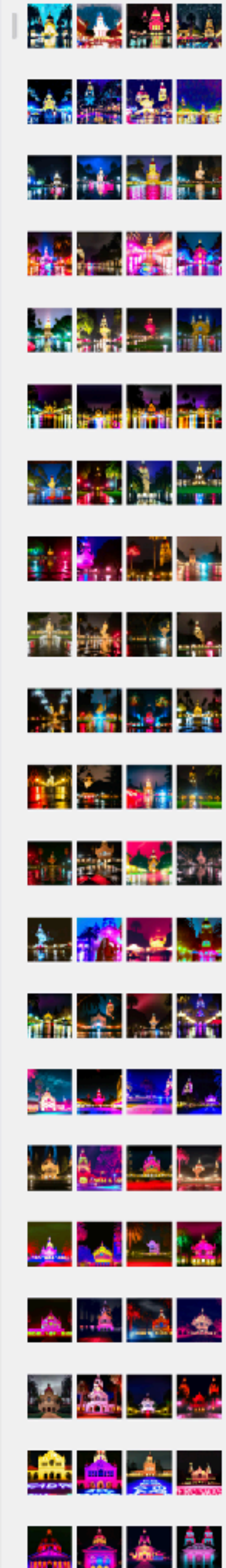
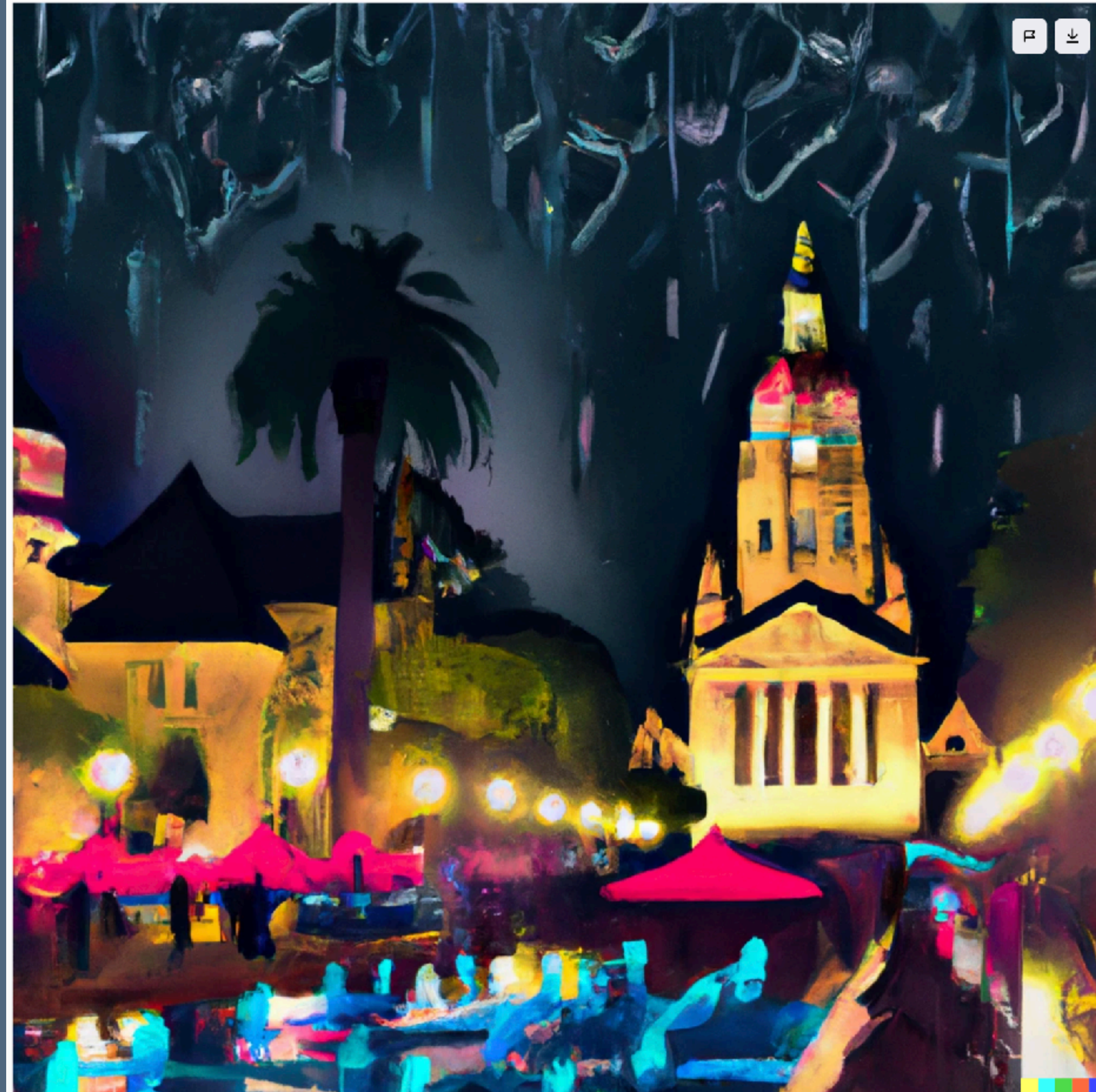


Iteration 8

nighttime rain stanford  
memorial church and  
main quad with palm  
trees, night market  
food stalls and neon  
signs **like downtown  
tokyo**



Iteration 17

[← Back](#)[Edit](#)[Variations](#)[Share](#)[Save](#)[→](#)

nighttime rain stanford  
memorial church and  
main quad with palm  
trees, night market  
**japadog** food stalls  
and neon signs, **neo**  
tokyo **bladerunner**  
style **film still**  
**illustration**

Iteration 21



**Aaron Hertzmann**

@AaronHertzmann



Writing a letter and quite happy with this phrase: Real artistic tools should act as extensions of the artist, the way a paintbrush adds capabilities to a painter's hand, rather than a slot machine that may or may not give you something useful.

8:05 AM · Sep 25, 2023 · **5,562** Views

Part of your job, as a designer, is to use the tools you have to create the best interfaces you can.

**Mitchell's take:** unpredictable, yet amazingly capable, black boxes can be incredible user interfaces... compared to what was possible before them.

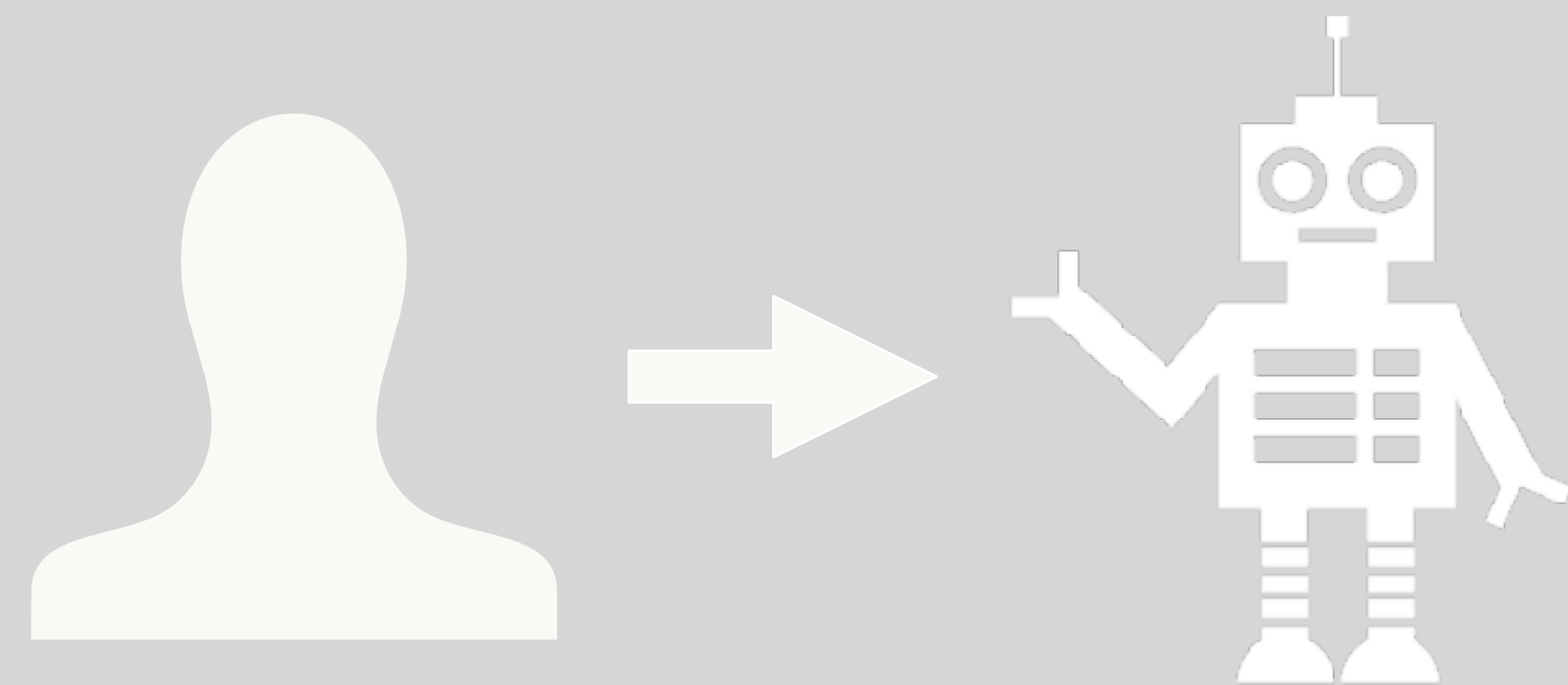
**How do you design  
ideal human-AI  
interactions?**

# **Intelligence Augmentation**

A reaction to:

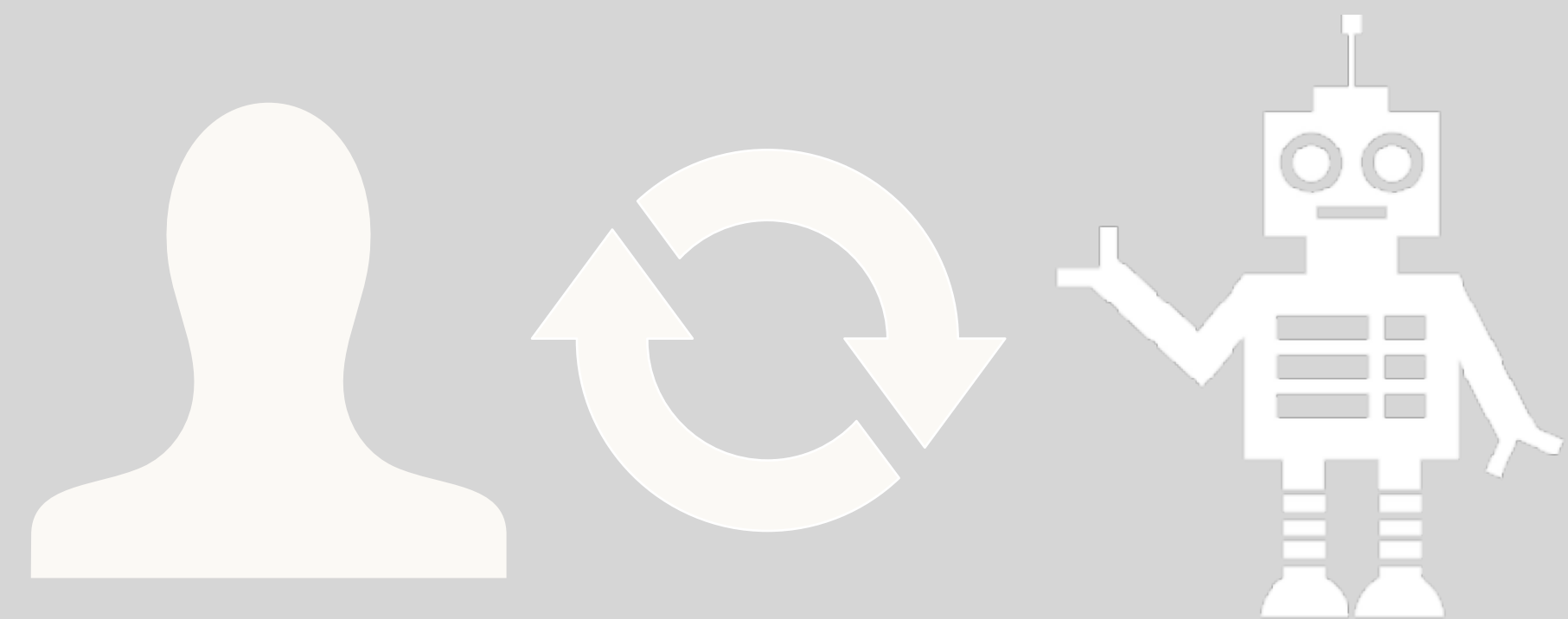
**“AI will replace human  
intelligence”**

# Artificial



Replace human intelligence  
with artificial intelligence

# Intelligence



Augment human intelligence  
with artificial intelligence

# Over half of LLM usage is intelligence augmentation today

Method: analyze requests sent to Anthropic's Claude system

## Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations

Kunal Handa\*, Alex Tamkin\*, Miles McCain, Saffron Huang, Esin Durmus

Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax, Kevin K. Troy

Dario Amodei, Jared Kaplan, Jack Clark, Deep Ganguli

Anthropic

### Abstract

Despite widespread speculation about artificial intelligence's impact on the future of work, we lack systematic empirical evidence about how these systems are actually being used for different tasks. Here, we present a novel framework for measuring AI usage patterns across the economy. We leverage a recent privacy-preserving system [Tamkin et al., 2024] to analyze over four million Claude.ai conversations through the lens of tasks and occupations in the U.S. Department of Labor's O\*NET Database. Our analysis reveals that AI usage primarily concentrates in software development and writing tasks, which together account for nearly half of all total usage. However, usage of AI extends more broadly across the economy, with  $\sim 36\%$  of occupations using AI for at least a quarter of their associated tasks. We also analyze *how* AI is being used for tasks, finding 57% of usage suggests augmentation of human capabilities (e.g., learning or iterating on an output) while 43% suggests automation (e.g., fulfilling a request with minimal human involvement). While our data and methods face important limitations and only paint a picture of AI usage on a single platform, they provide an automated, granular approach for tracking AI's evolving role in the economy and identifying leading indicators of future impact as these technologies continue to advance.

# INTRODUCTION

\*\*\*\*\*

OVERALL ABOUT PROGRAM  
SEE AS AN \*INSTRUMENTA  
CONTROL TECHNIQUES  
SEE IMPLEMENTATION  
USAGE  
ACTIVITIES  
CREDITS



# AUGMENTING HUMAN INTELLECT: A CONCEPTUAL FRAMEWORK

*Prepared for:*

DIRECTOR OF INFORMATION SCIENCES  
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH  
WASHINGTON 25, D.C.

CONTRACT AF 49(638)-1024

*By: D. C. Engelbart*

STANFORD RESEARCH INSTITUTE

MENLO PARK, CALIFORNIA



# AUGMENTING HUMAN INTELLECT

## I INTRODUCTION

### A. GENERAL

By "augmenting human intellect" we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems. Increased capability in this respect is taken to mean a mixture of the following: more-rapid comprehension, better comprehension, the possibility of gaining a useful degree of comprehension in a situation that previously was too complex, speedier solutions, better solutions, and the possibility of finding solutions to problems that before seemed insoluble. And by "complex situations" we include the professional problems of diplomats, executives, social scientists, life scientists, physical scientists, attorneys, designers--whether the problem situation exists for twenty minutes or twenty years. We do not speak of isolated clever tricks that help in particular situations. We refer to a way of life in an integrated domain where hunches, cut-and-try, intangibles, and the human "feel for a situation" usefully co-exist with powerful concepts, streamlined terminology and notation, sophisticated methods, and high-powered electronic aids.

**Why augment  
instead of  
replace?**

## Abstract

Big Data evangelists often argue that algorithms make decision-making more informed and objective—a promise hotly contested by critics of these technologies. Yet, to date, most of the debate has focused on the instruments themselves, rather than on how they are used. This article addresses this lack by examining the actual *practices* surrounding algorithmic technologies. Specifically, drawing on multi-sited ethnographic data, I compare how algorithms are used and interpreted in two institutional contexts with markedly different characteristics: web journalism and criminal justice. I find that there are surprising similarities in how web journalists and legal professionals use algorithms in their work. In both cases, I document a gap between the intended and actual effects of algorithms—a process I analyze as “decoupling.” Second, I identify a gamut of buffering strategies used by both web journalists and legal professionals to minimize the impact of algorithms in their daily work. Those include foot-dragging, gaming, and open critique. Of course, these similarities do not exhaust the differences between the two cases, which are explored in the discussion section. I conclude with a call for further ethnographic work on algorithms in practice as an important empirical check against the dominant rhetoric of algorithmic power.

## Keywords

Algorithms, ethnography, work practices, organizations, journalism, criminal justice

## Introduction

We live in an era of data: an unprecedented amount of digital information is being collected, stored, and ana-

of using “smart statistics” to “disrupt” or “moneyball” sectors with long histories of inefficiency and bias (Castro, 2016; Milgram, 2013). On the other hand, scholars criticize the “mythology” of Big Data (boyd

If you try  
thoughtlessly...



[News](#) › [Stories](#) › [Archives](#) › [2019](#) › [May](#) › CMU Researchers Make Transformational AI Seem "Unremarkable"

May 08, 2019

## CMU Researchers Make Transformational AI Seem "Unremarkable"

AI must be unobtrusive to be accepted as part of clinical decision making

# Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes

Qian Yang  
HCI Institute  
Carnegie Mellon University  
yangqian@cmu.edu

Aaron Steinfeld  
Robotics Institute  
Carnegie Mellon University  
steinfeld@cmu.edu

John Zimmerman  
HCI Institute  
Carnegie Mellon University  
johnz@cs.cmu.edu

### ABSTRACT

Clinical decision support tools (DST) promise improved healthcare outcomes by offering data-driven insights. While effective in lab settings, almost all DSTs have failed in practice. Empirical research diagnosed poor contextual fit as the cause. This paper describes the design and field evaluation of a radically new form of DST. It automatically generates slides for clinicians' decision meetings with subtly embedded machine prognostics. This design took inspiration from the notion of *Unremarkable Computing*, that by augmenting the users' routines technology/AI can have significant importance for the users yet remain unobtrusive. Our field evaluation suggests clinicians are more likely to encounter and embrace such a DST. Drawing on their responses, we discuss the importance and intricacies of finding the right level of unremarkableness in DST design, and share lessons learned in prototyping critical AI systems as a situated experience.

### CCS CONCEPTS

• **Human-centered computing** → *User centered design*;

### KEYWORDS

Decision Support Systems, Healthcare, User Experience.

### ACM Reference Format:

Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, May 4–9, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3290605.3300468>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*CHI 2019*, May 4–9, 2019, Glasgow, Scotland Uk  
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-5970-2/19/05...\$15.00  
<https://doi.org/10.1145/3290605.3300468>

### 1 INTRODUCTION

The idea of leveraging machine intelligence in healthcare in the form of decision support tools (DSTs) has fascinated healthcare and AI researchers for decades. These tools often promise insights on patient diagnosis, treatment options, and likely prognosis. With the adoption of electronic medical records and the explosive technical advances in machine learning (ML) in recent years, now seems a perfect time for DSTs to impact healthcare practice.

Interestingly, almost all these tools have failed when migrating from research labs to clinical practice in the past 30 years [5, 8, 9]. In a review of deployed DSTs, healthcare researchers ranked the lack of HCI considerations as the most likely reason for failure [12, 23]. This includes a lack of consideration for clinicians' workflow and the collaborative nature of clinical work. The interaction design of most clinical decision support tools instead assumes that individual clinicians will recognize when they need help, walk up and use a system that is separate from the electronic health record, and that they want and will trust the system's output.


We are collaborating with biomedical researchers on the design of a DST supporting the decision to implant an artificial heart. The artificial heart, VAD (ventricular assist device), is an implantable electro-mechanical device used to partially replace heart function. For many end-stage heart failure patients who are not eligible for or able to receive a heart transplant, VADs offer the only chance to extend their lives. Unfortunately, many patients who received VADs die shortly after the implant [2]. In this light, a DST that can predict the likely trajectory a patient will take post-implant, should help identify the patients who are mostly likely to benefit from the therapy.

We draw insight from a field study investigating the VAD decision processes, searching for opportunities where ML might help [26]. The findings revealed that clinicians are unlikely to encounter or to actively engage with a DST for help at the time and place of decision making. For most cases, they did not find the implant decision challenging; thus, they had no desire for computational support. In addition, the extremely hierarchical healthcare culture stratified senior physicians who make implant decisions and the


**Goal:**


**human+AI > human**

We call this “complementarity”



Stanford University  
Human-Centered  
Artificial Intelligence


Search this site


Open

Design and Human-Computer Interaction, Economy and Markets


# Will Generative AI Make You More Productive at Work? Yes, But Only If You're Not Already Great at Your Job.

Scholars examining the impact of an AI assistant at a call center find gains for less experienced workers. ...

BCG

Log in


- Around 90% of participants improved their performance when using GenAI for creative ideation. People did best when they did not attempt to edit GPT-4's output.
- When working on business problem solving, a task outside the tool's current competence, many participants took GPT-4's misleading output at face value. Their performance was 23% worse than those who didn't use the tool at all.



Eric Topol
@EricTopol


The largest medical [#AI](#) randomized controlled trial yet performed, enrolling >100,000 women undergoing mammography screening, was published today [@LancetDigitalH](#)

The use of A.I. led to 29% higher detection of cancer, no increase of false positives, and reduced workload compared with radiologists without A.I.. [thelancet.com/journals/landi...](https://thelancet.com/journals/landi...)



Elizabeth Barnes
@BethMayBarnes

Our RCT found that [early-2025] AI coding assistants appear to \*slow down\* users [working in mature open-source codebases]. But developer self-reports (and expert forecasts) suggested speedup. This is a counterintuitive result! Some thoughts on interpretations / takeaways



METR
@METR\_Evals · Jul 10

We ran a randomized controlled trial to see how much AI coding tools speed up experienced open-source developers.

The results surprised us: Developers thought they were 20% faster with AI tools, but they were actually 19% slower when they had access to AI than ...

### Against Expert Forecasts and Developer Self-Reports, Early-2025 AI Slows Down Experienced Open-Source Developers

In this RCT, 16 developers with moderate AI experience complete 246 tasks in large and complex projects on which they have an average of 5 years of prior experience.

Analysis of 106 studies covering 370 effect sizes

On average, human-AI combinations **perform worse** than the best of humans or AI alone

Biggest **losses** for decision-making tasks and biggest **wins** for content creation tasks

## nature human behaviour

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

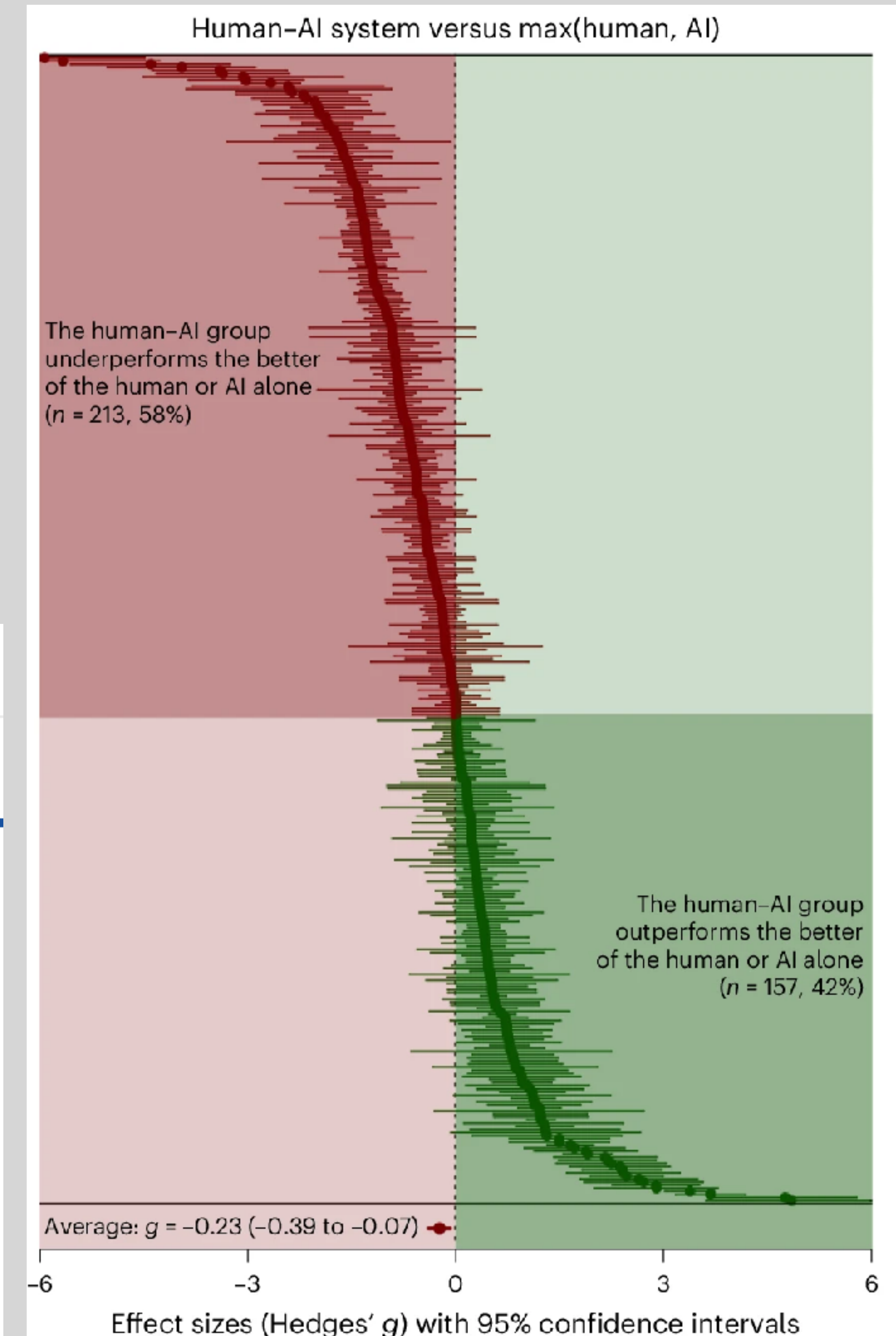
[nature](#) > [nature human behaviour](#) > [articles](#) > article

Article | [Open access](#) | Published: 28 October 2024

# When combinations of humans and AI are useful: A systematic review and meta-analysis

[Michelle Vaccaro](#), [Abdullah Almaatouq](#) & [Thomas Malone](#) 

[Nature Human Behaviour](#) (2024) | [Cite this article](#)



**Goal:**

**human+AI  $\approx$  human**

We call this “not great”

# Why?

**Overreliance:** When the algorithm suggests the answer to you, you get influenced by the AI's suggestion and rely on it when we shouldn't [Buçinca, Malaya, and Gajos 2021]

...even if the algorithm explains its reasoning, unless the explanation takes almost no effort to verify [Vasconcelos et al. 2023]

**Algorithm aversion:** we prefer human decision-making to AIs, even if the algorithm is better at the task [Dietvorst, Simmons, and Massey 2015]

...and especially after seeing the algorithm make an error

# How to Achieve Intelligence Augmentation

- **Look for gaps:** keenly felt gaps in information, knowledge, or execution.
  - e.g.: “How might [this group] react to [this message]?”, “What is a concise summary of the project status?”, “This situation is turning into a conflict. What might happen if I [take this action]?”
- If you fill the gap, it enables me to be better at what I do
  - Curtis Langlotz: **“AI won't replace radiologists, but radiologists who use AI will replace those who don't.”**

# How do we rapidly prototype AI solutions?

**Prompt prototyping:** Use ChatGPT—give it an example input to your problem, and tell it what you want it to do. Does it do roughly the right thing? If it's close, then you've got a good bet.

Please determine whether this forum comment in the Cisco Webex customer support "civil" or "incivil"?

Here is the comment: "Go shut yourself in your room and think about what you just wrote."

# “But how do we convince the board?”

The usual narrative: we know how to convey the value of AI through savings. But how do we convince people about the value of augmentation?

**What's your goal?** Replacement is about reducing costs. But augmentation might be increasing performance, reducing errors, or making more effective decisions. Align your metrics with your goals

# Summary

- Modern AI models open new opportunities for product development
- However, even the smartest AI based product ultimately needs to solve a real problem for real people
- Use intelligence augmentation as your litmus test