Human-Centered Artificial Intelligence

CS 347

Michael Bernstein

Last time

Artifacts have politics: the systems we create influence groups and societies, often with undesirable outcomes

> **Example: gig economy** — potential of upward mobility and community social capital, but not currently implemented in a way that unlocks those possibilities
>
> Design approaches focused on marginalized groups, such as **feminist HCI**, center these communities' needs in the design process
>
> **Algorithmic systems**, not just designed systems, similarly have impact. People struggle to reason about them, and industry struggles to avoid mistakes. But, **modeling human-centered objectives** can help.

Social Computing

Unit 3

social media
collaboration
design + society

Where we go from here

so far	Ubiquitous Computing, Design, Social Computing
week 5	Human-Centered AI
week 6	Toolkits/Content Creation
week 7	Visualization/Cognitive Models
week 8	Health/Education/Critical Theory/HCI community
week 9	Methodology/Accessibility
week 10	ICT4D/Something Old, Something New

Software and Tools

Unit 4

human-centered AI
tools and toolkits
content creation

Today

AI vs. IA

Direct manipulation vs. Agents

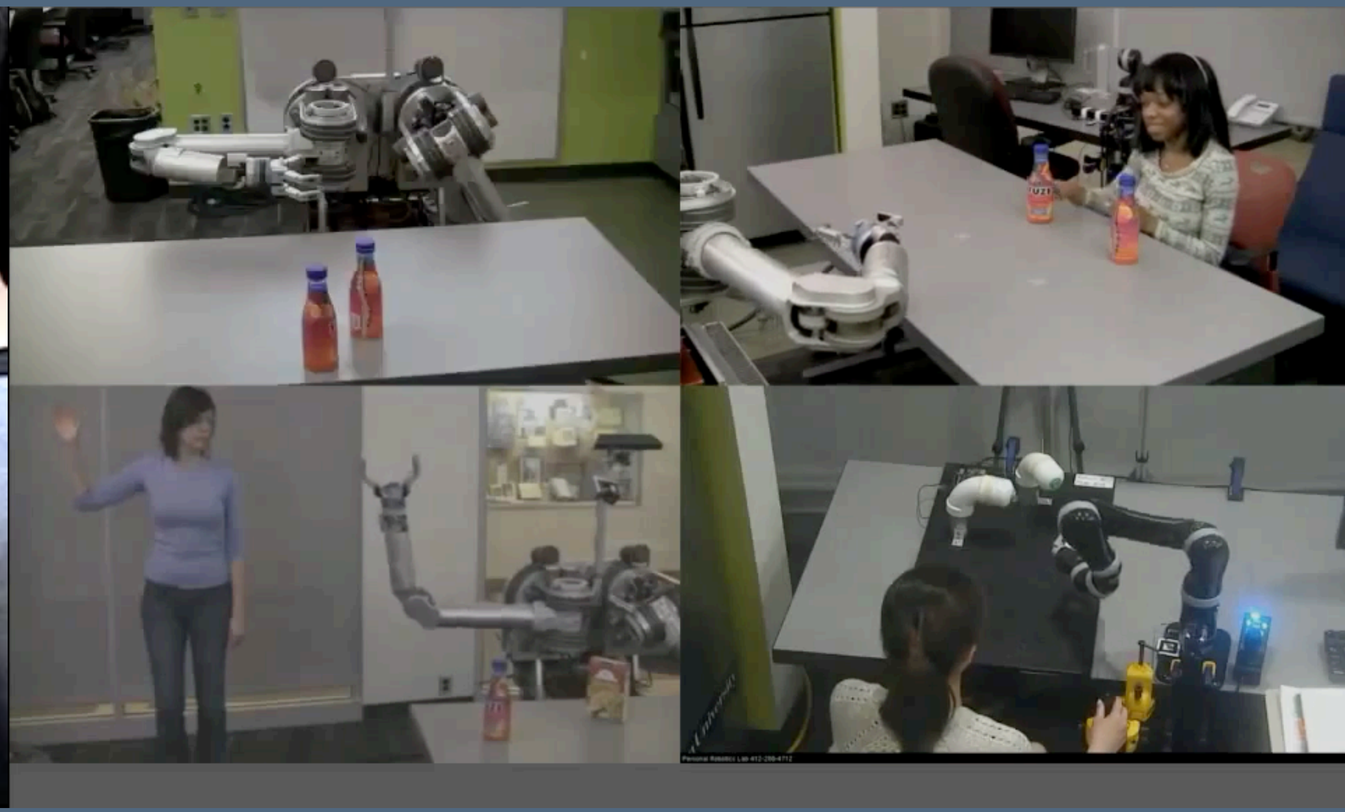
Mixed-initiative interaction

End-user AI authoring

AI and design

People: where AI lives or dies

[Breazeal 2004]



[Dragan, Lee, and Srinivasa 2013]

…but we need to think carefully

[Mok et al. 2015]

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

- Eytan Adar [2018]

Intelligence Augmentation

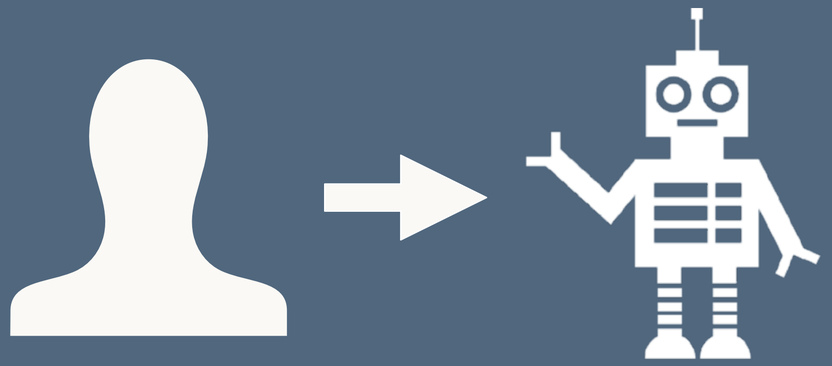
AI vs. IA

A reaction to:

“AI will replace human intelligence”

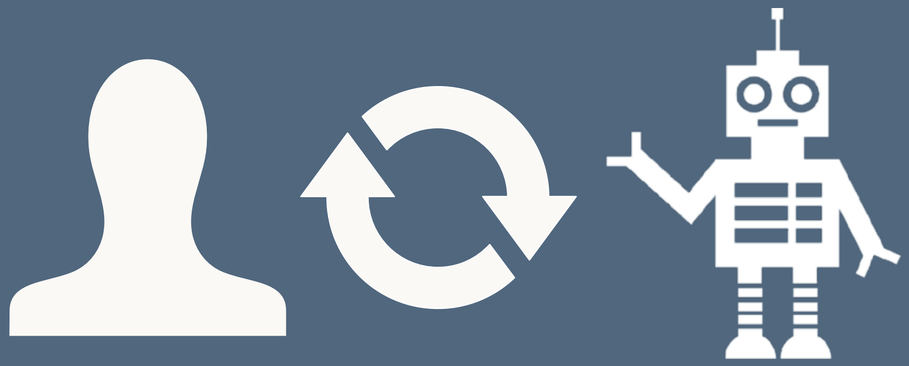
Intelligence augmentation says that **replacement is the wrong approach**.

Artificial Intelligence



Replace human intelligence with artificial intelligence

Intelligence Augmentation



Augment human intelligence with artificial intelligence

Big Data & Society
July–December 2017: 1–14

DOI: 10.1177/2053951717718855
journals.sagepub.com/home/bds

Algorithms in practice: Comparing web journalism and criminal justice

Angèle Christin

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.

If you try thoughtlessly…

Keywords

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

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

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

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”

Artificial Intelligence, Scientific Discovery, and Product Innovation*

Aidan Toner-Rodgers†
MIT

November 6, 2024

This paper studies the impact of artificial intelligence on innovation, exploiting the randomized introduction of a new materials discovery technology to 1,018 scientists in the R&D lab of a large U.S. firm. AI-assisted researchers discover 44% more materials, resulting in a 39% increase in patent filings and a 17% rise in downstream product innovation. These compounds possess more novel chemical structures and lead to more radical inventions. However, the technology has strikingly disparate effects across the productivity distribution: while the bottom third of scientists see little benefit, the output of top researchers nearly doubles. Investigating the mechanisms behind these results, I show that AI automates 57% of "idea-generation" tasks, reallocating researchers to the new task of evaluating model-produced candidate materials. Top researchers leverage their domain knowledge to prioritize promising AI suggestions, filtering out false positives. Together, these findings and highlight the complemen- Survey evidence

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...

BCG

READING TIME: 15 MIN

Log in

Key Takeaways

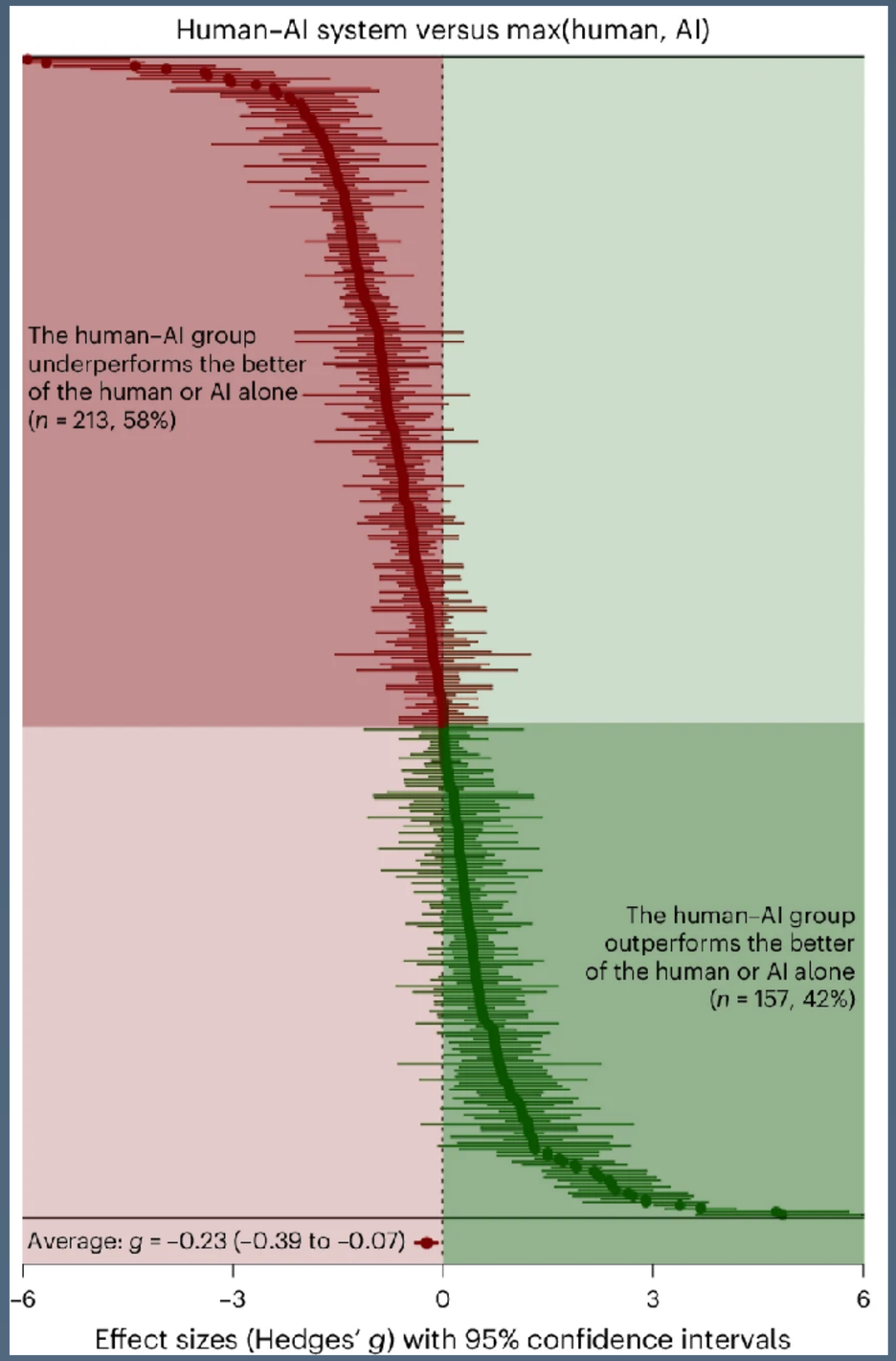
A first-of-its-kind scientific experiment finds that people mistrust generative AI in areas where it can contribute tremendous value and trust it too much where the technology isn't competent.

- 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.

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

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 > human

We call this “complementarity”

Goal:
human+AI ≈ human

We call this “not great”

Introduction

Overall about Program

NLS as an "Instrument"

Control techniques

NLS Implementation

Usage

Activities

Credits

Engelbart Institute

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

*SRI

Augmentation examples we've discussed

Help me understand where I'm using water in my household

Realize my sketched mechanical design into a rough functional system

Connect me with jobs or movies that I might want to see

Show me behavior patterns that are influencing my health

But who should lead this dance? How much control should we yield to the AI? This leads to a debate…

YOU READ THIS

Agents vs. Direct Manipulation

[Shneiderman and Maes 1997]

Software agents

We should delegate to proactive artificial intelligence systems

Pattie Maes, MIT Media Lab

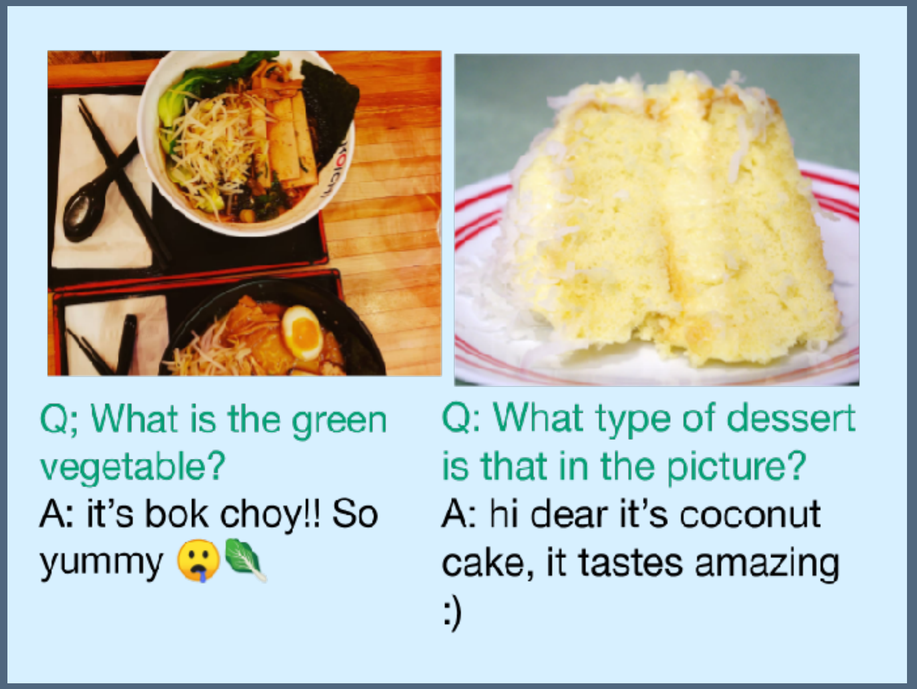
Direct manipulation

Users should always have full control, even as automation increases

Ben Shneiderman, U. Maryland

Agents

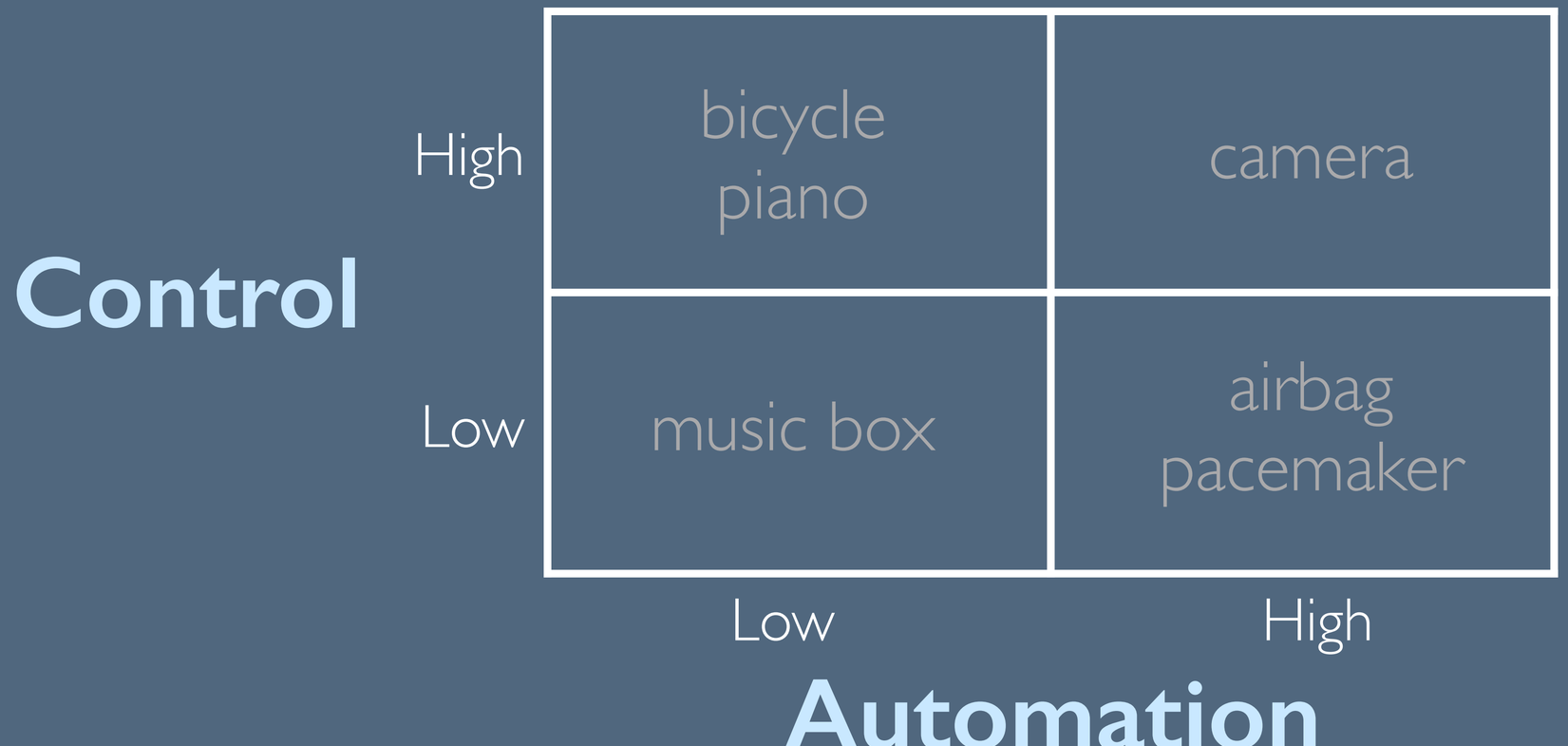
AI agents ask questions about images on social media to learn about the world around them [Krishna et al. 2022]



Learn to automate tasks that you do commonly [Maes 1995]

Direct manipulation

Shneiderman: it is possible to maintain high levels of user control even as automation increases [Shneiderman 2022]



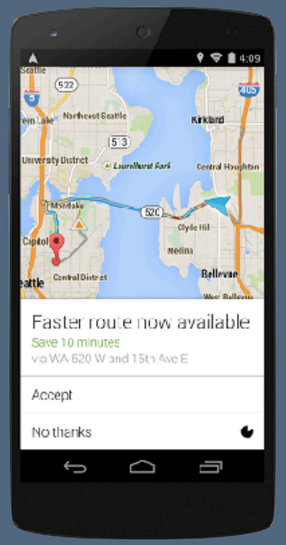
Mixed initiative interaction

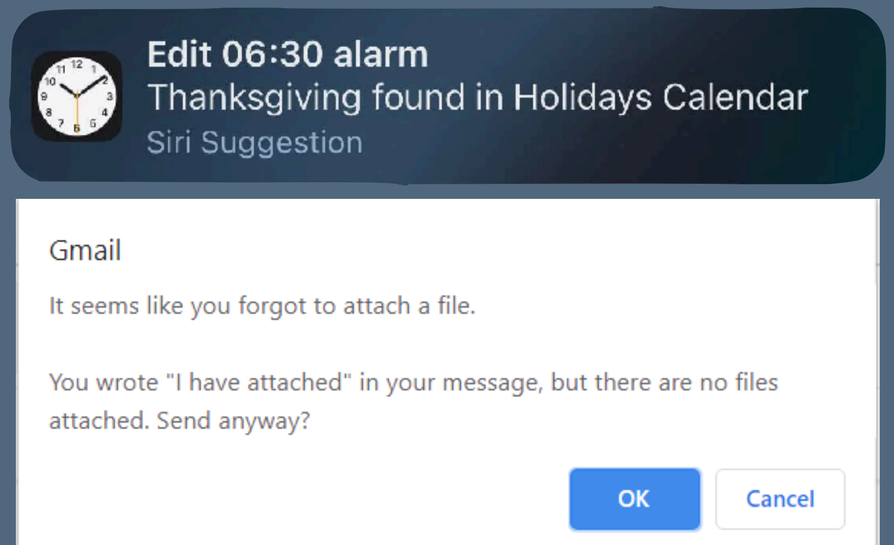
Eric Horvitz keeps listening to the agents vs. direct manipulation debate. He decides that he's had enough and that it's a false dichotomy…

Mixed-initiative, intuitively

You don't need to decide between full control and full automation. Instead, the system should automate the things it can, hand control to the user for the things it can't, and ask the user if it's unsure.

Today, mixed-initiative interaction typically refers to the mode of suggesting an action and letting the user confirm it





Mixed-initiative as utilities

[Horvitz 1999]

Horvitz envisioned mixed-initiative more broadly as trading off dynamically between all options, using **utilities**:

Numbers representing the benefit or harm of an outcome

u(A,G) = (positive) utility of taking an automated action when the goal is correctly guessed

u(A,¬G) = (negative) utility of taking the same action when the goal is incorrectly guessed

u(¬A,G) and u(¬A,¬G) similarly

	Desired goal	Not desired goal
Take action	u(A,G)	u(A,¬G)
No action	u(¬A,G)	u(¬A,¬G)

Now, take expected values

[Horvitz 1999]

What's the expected value of taking action?

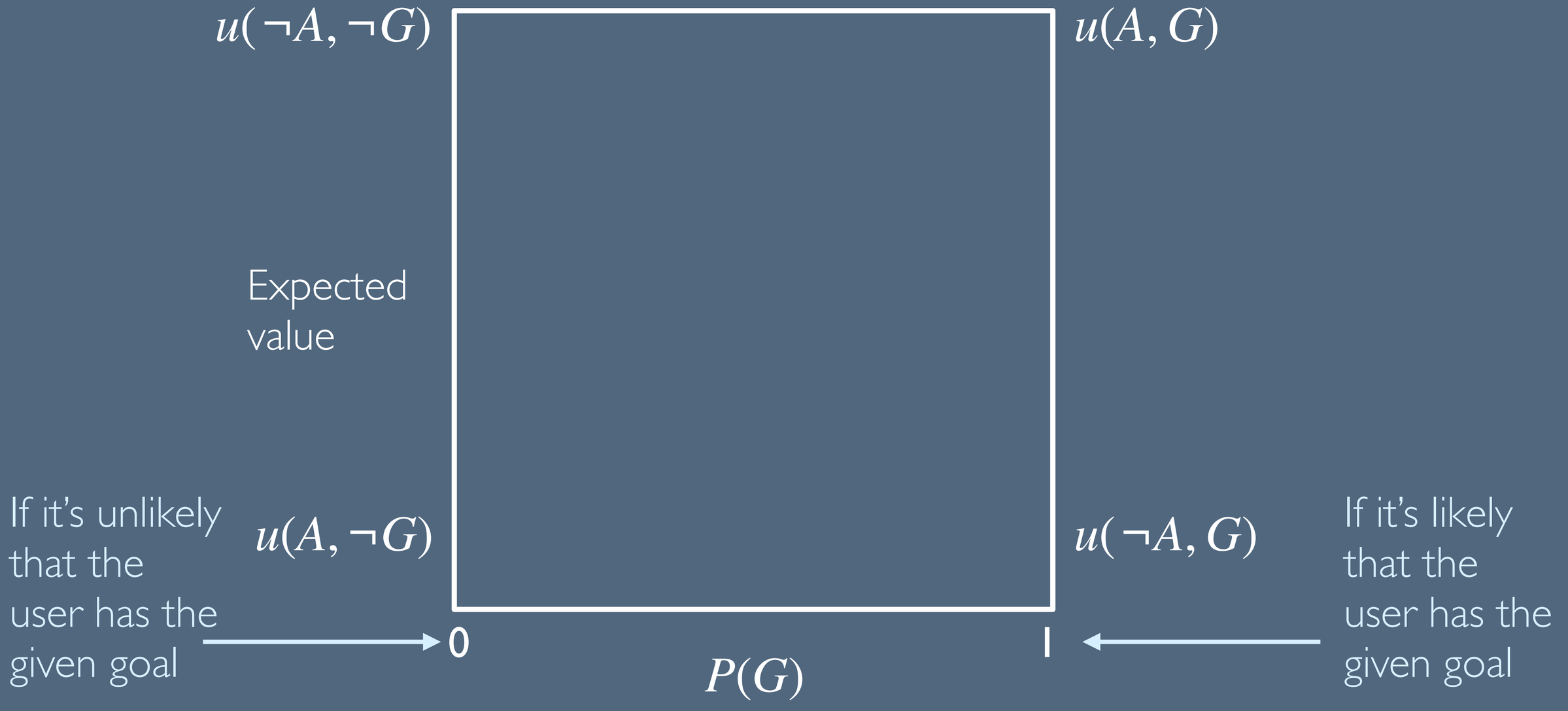
$$P(G) \cdot u(A, G) + P(\neg G) \cdot u(A, \neg G)$$

What's the expected value of taking no action?

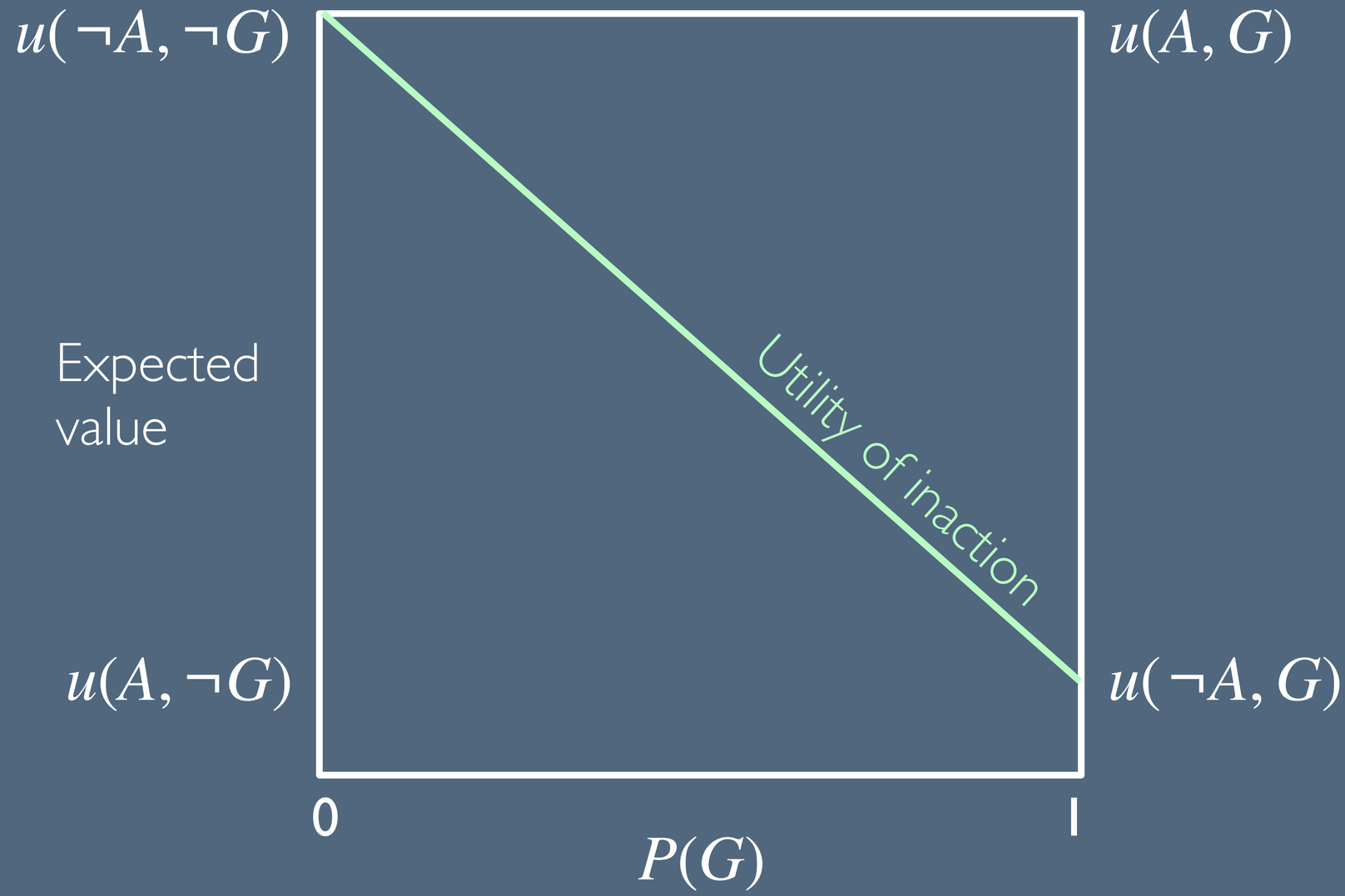
$$P(G) \cdot u(\neg A, G) + P(\neg G) \cdot u(\neg A, \neg G)$$

	Desired goal	Not desired goal
Take action	u(A,G)	u(A,¬G)
No action	u(¬A,G)	u(¬A,¬G)

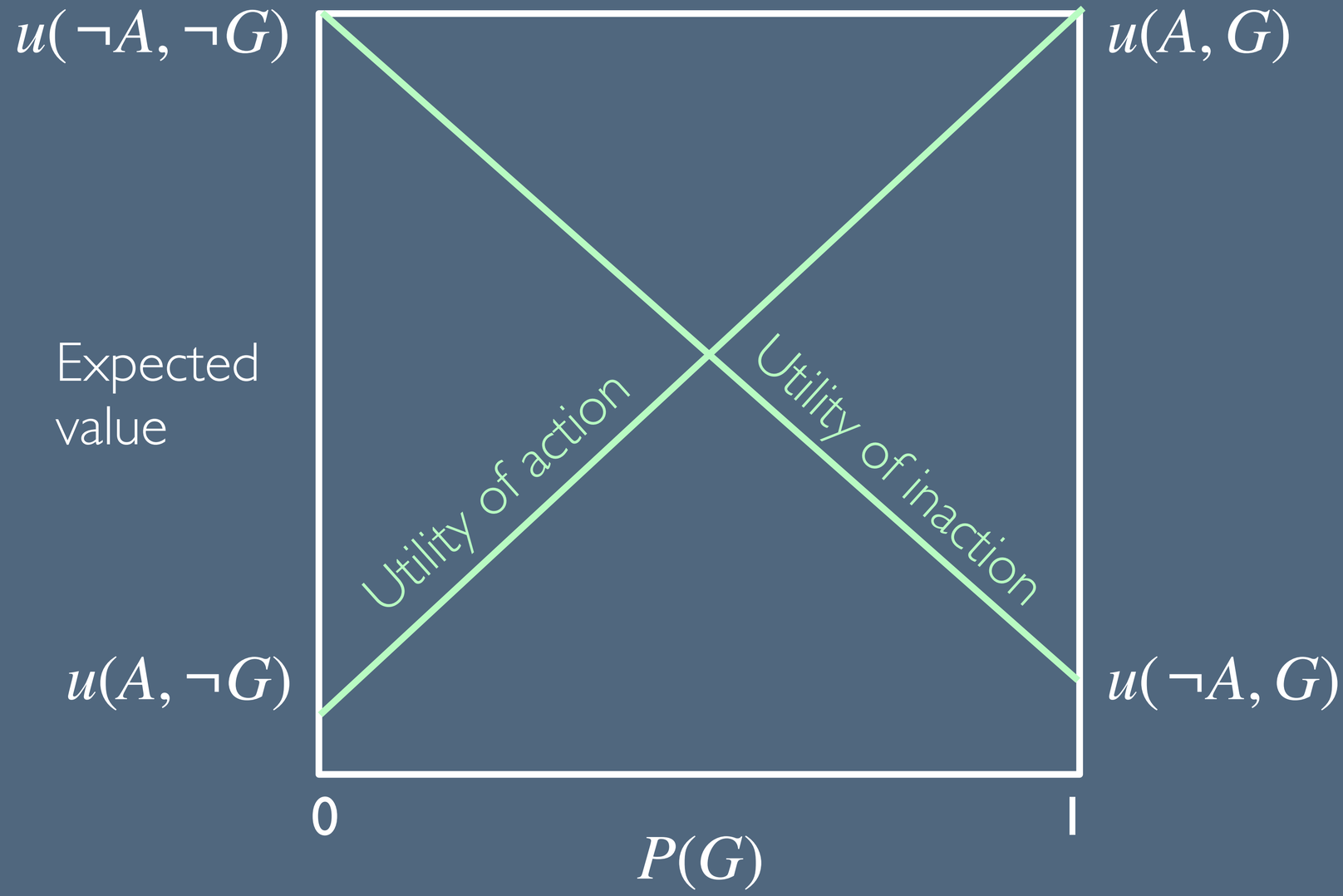
Mixed initiative: visually



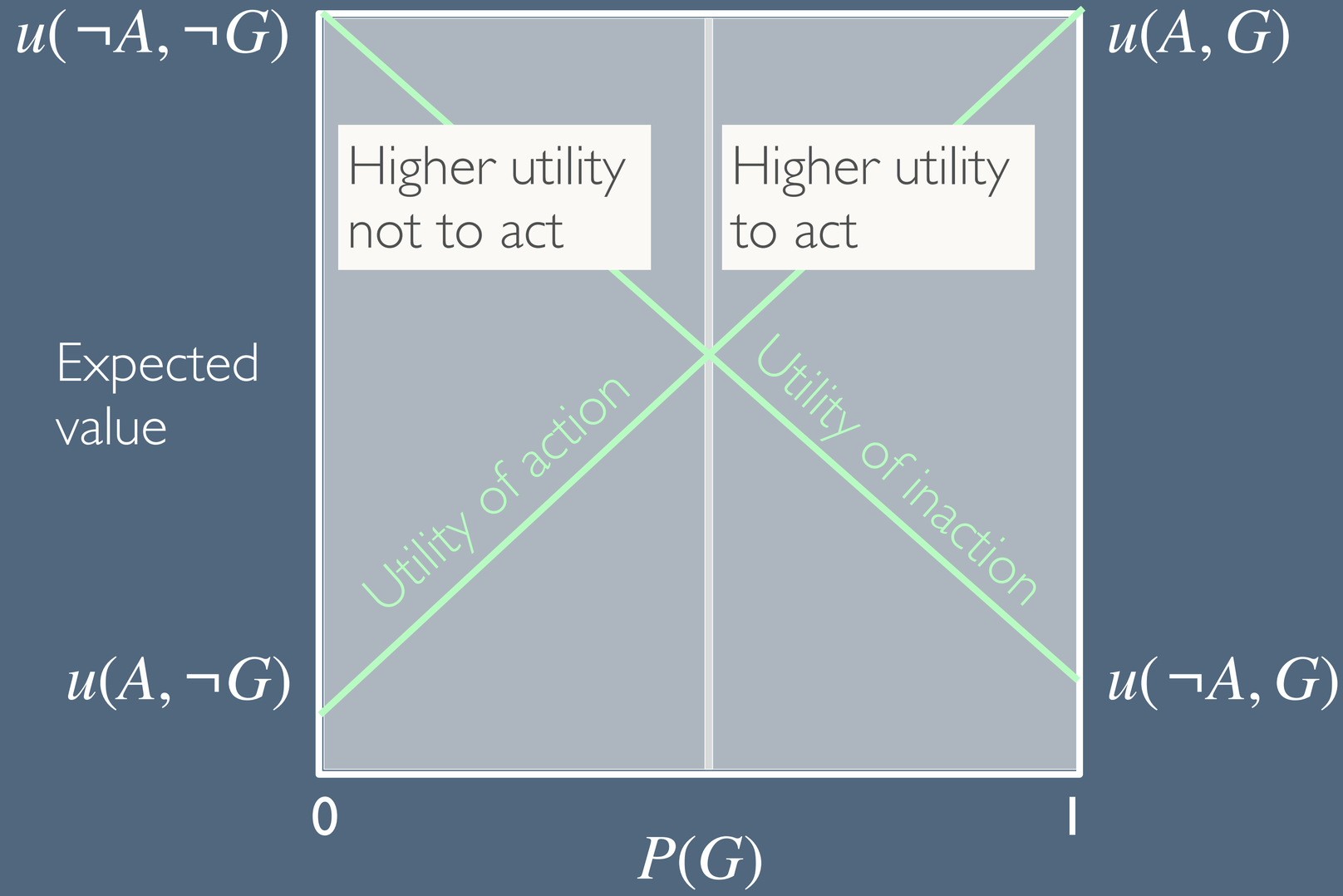
Mixed initiative: visually



Mixed initiative: visually

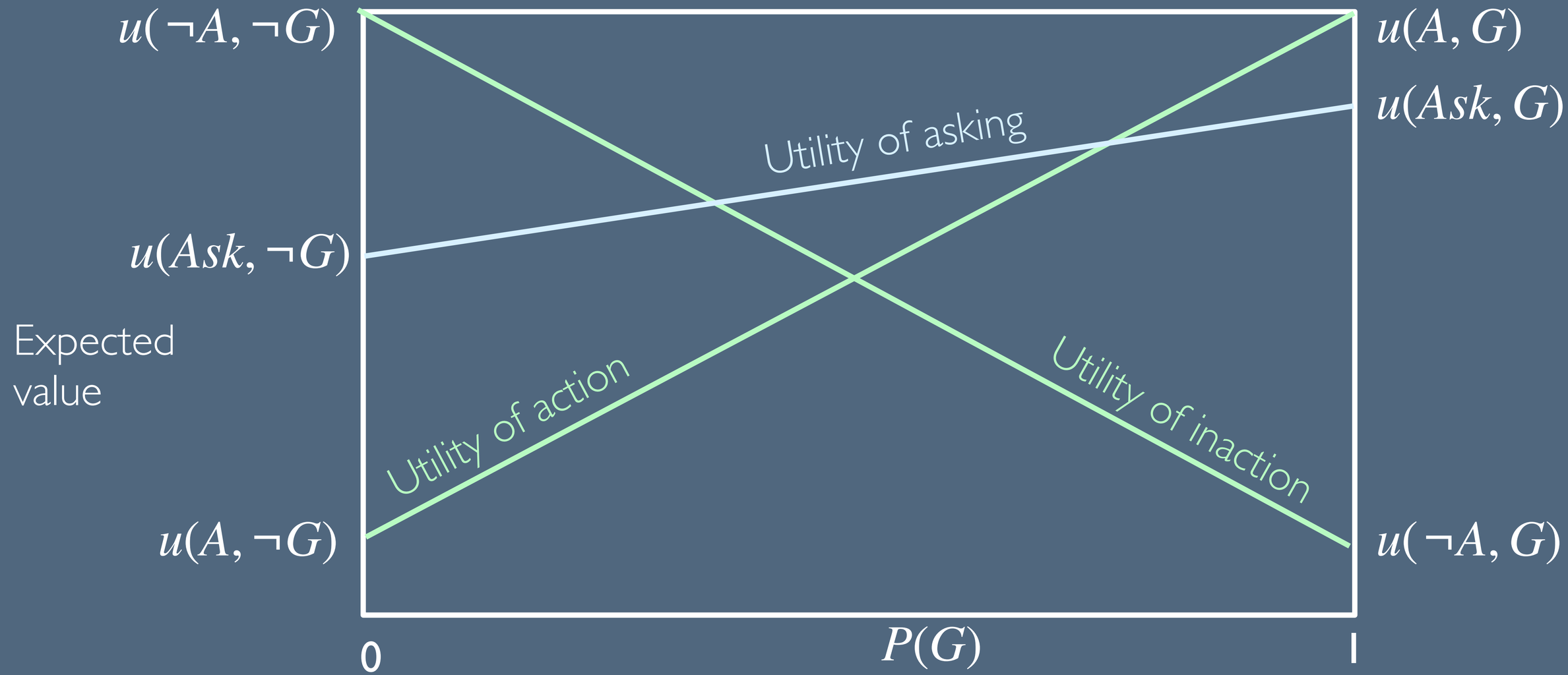


Mixed initiative: visually



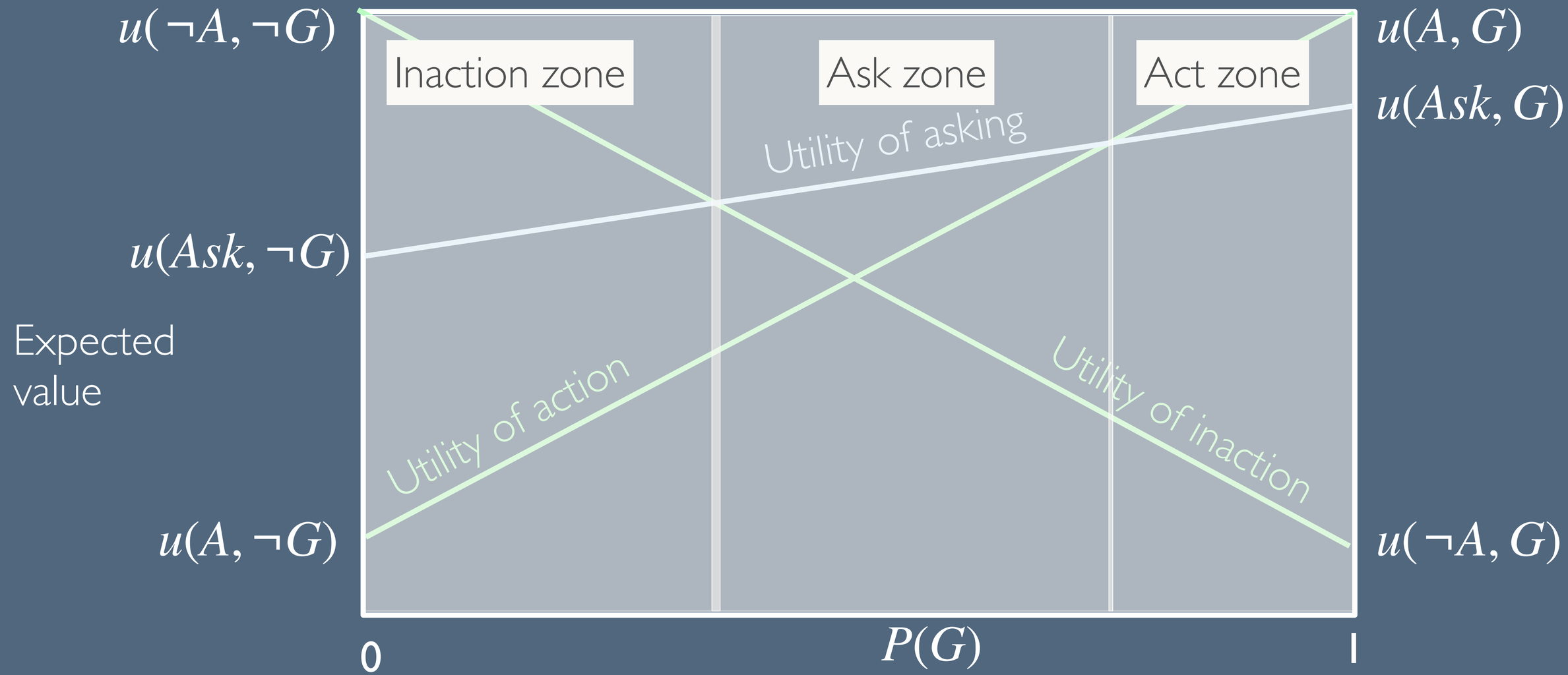
What if we ask the user?

Asking often carries lower risk, but also lower utility



What if we ask the user?

Asking often carries lower risk, but also lower utility



So, when does this screw up?

When the system cannot accurately assess the probability of the user having the goal P(G)

or

When the utilities are not correctly estimated

> e.g., too high a utility for asking if the user doesn't have the goal G. "Are you writing a letter right now?"

A problem has been detected and Windows has been shut down to prevent damage
to your computer.

The problem seems to be caused by the following file: kbdhid.sys

MANUALLY_INITIATED_CRASH

If this is the first time you've seen this stop error screen,
restart your computer. If this screen appears again, follow
these steps:

Check to make sure any new hardware or software is properly installed.
If this is a new installation, ask your hardware or software manufacturer
for any Windows updates you might need.

If problems continue, disable or remove any newly installed hardware
or software. Disable BIOS memory options such as caching or shadowing.
If you need to use safe mode to remove or disable components, restart
your computer, press F8 to select Advanced Startup Options, and then
select Safe Mode.

Technical Information:

*** STOP: 0x000000e2 (0x00000000, 0x00000000, 0x00000000, 0x00000000)

End user authoring of artificial intelligence

If you wanted a private smart doorbell...

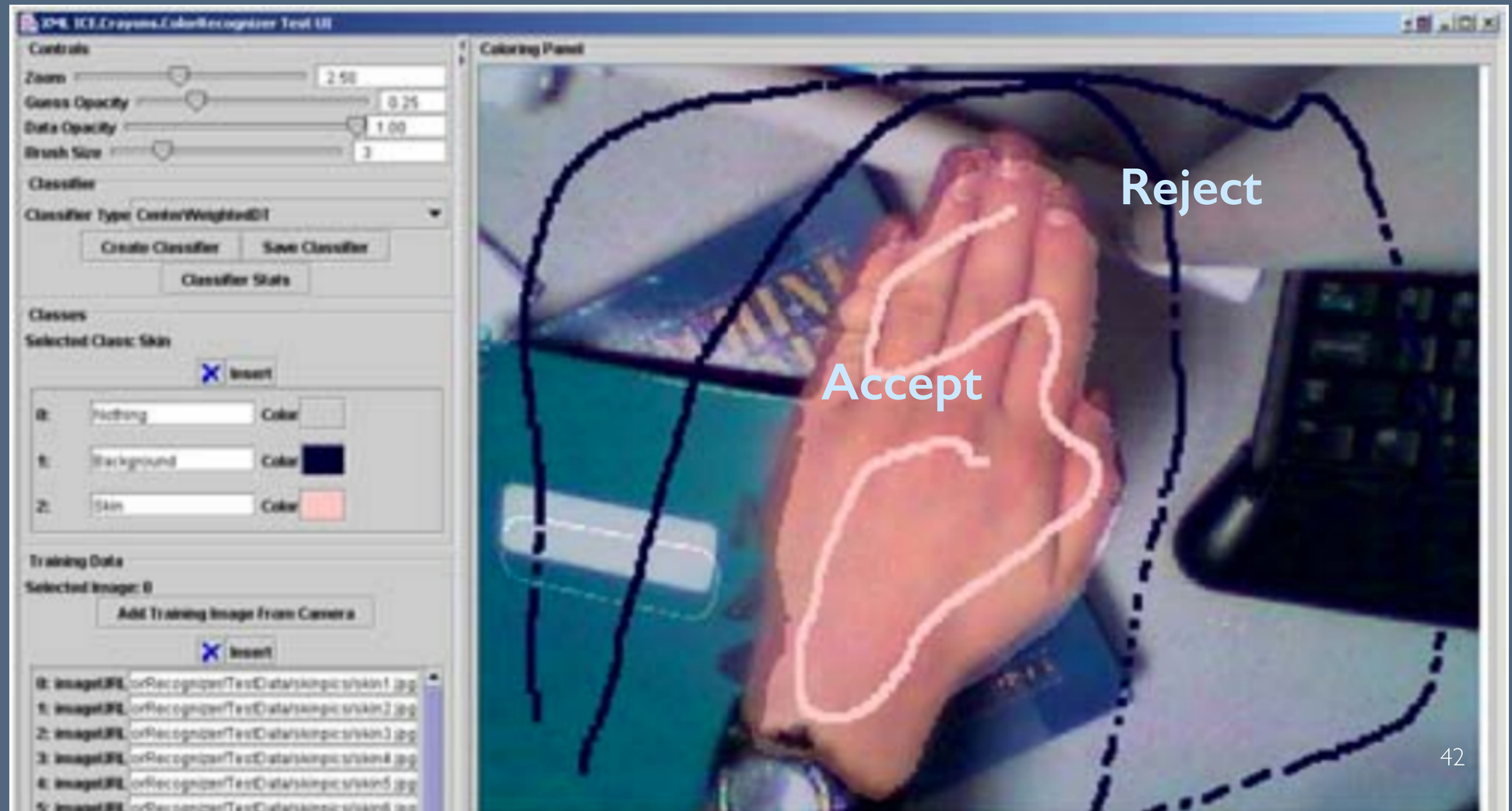
To automatically control entrance to your room to let in possible donors for your Stanford education

How might we let people train such a doorbell

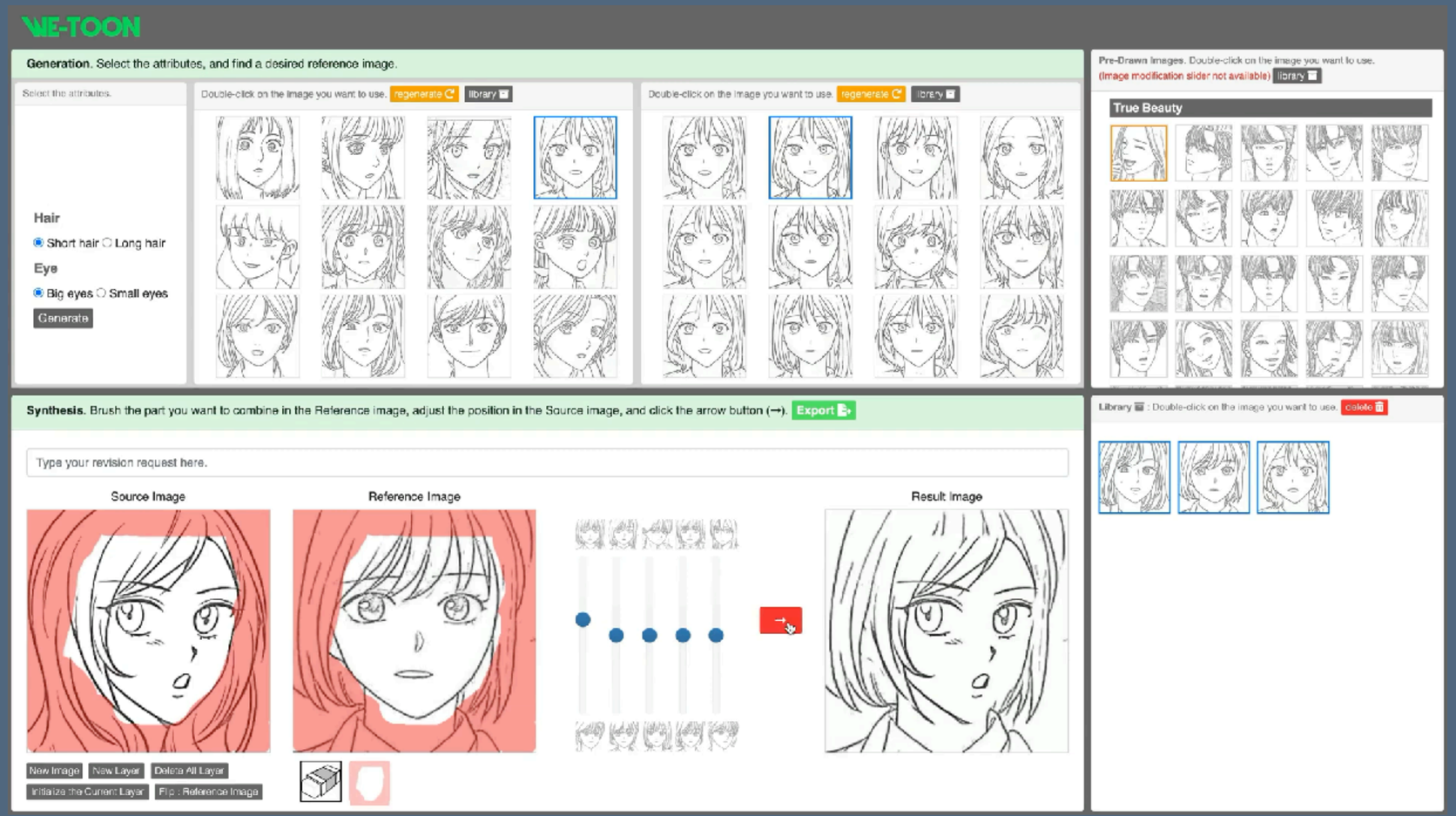
Crayons: camera-based interaction

[Fails and Olsen 2003]

"The one that started it all": direct-manipulation training



Frontier: image editing through demonstration

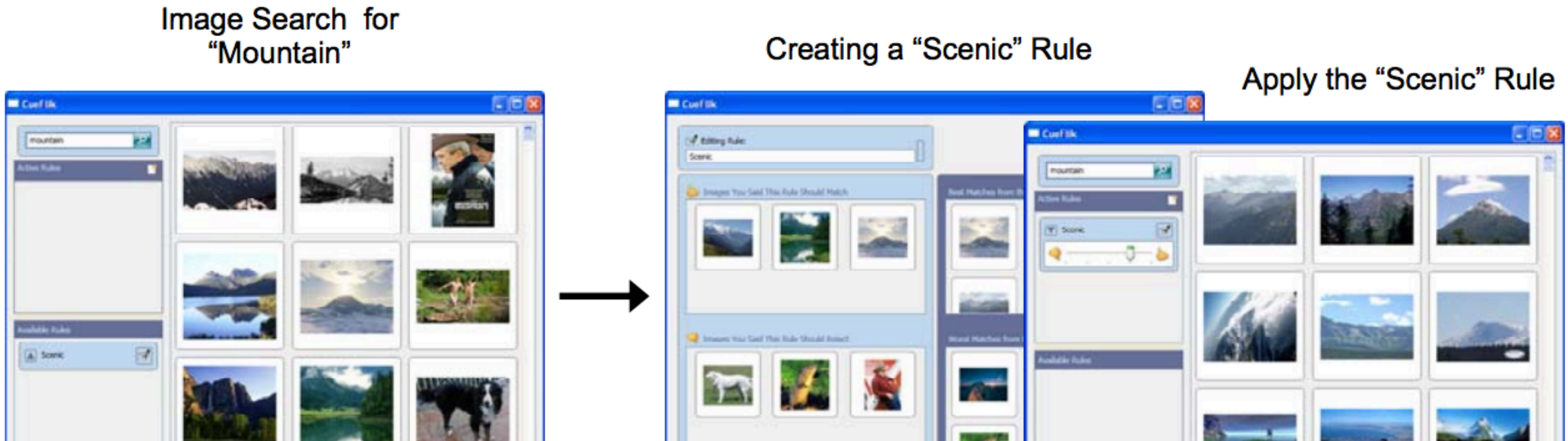


"Make this part of the source image look more like the reference image." [Ko et al. 2022]

Interactive training

[Fogarty et al. 2008]

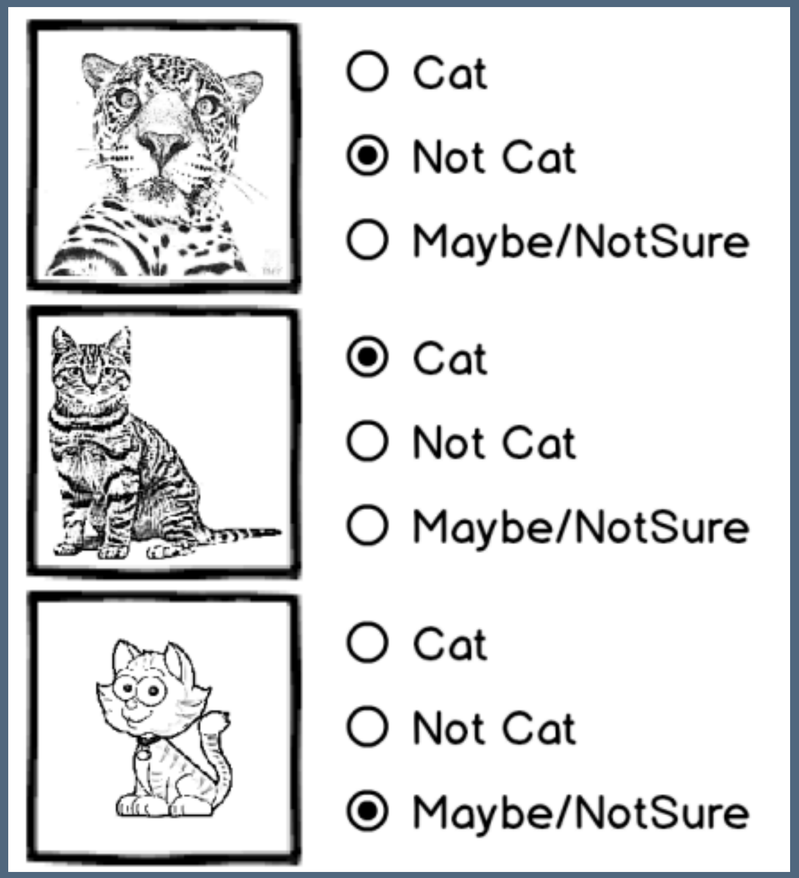
Allow users to keep training and re-training by drag-dropping instances into positive and negative classes as they go

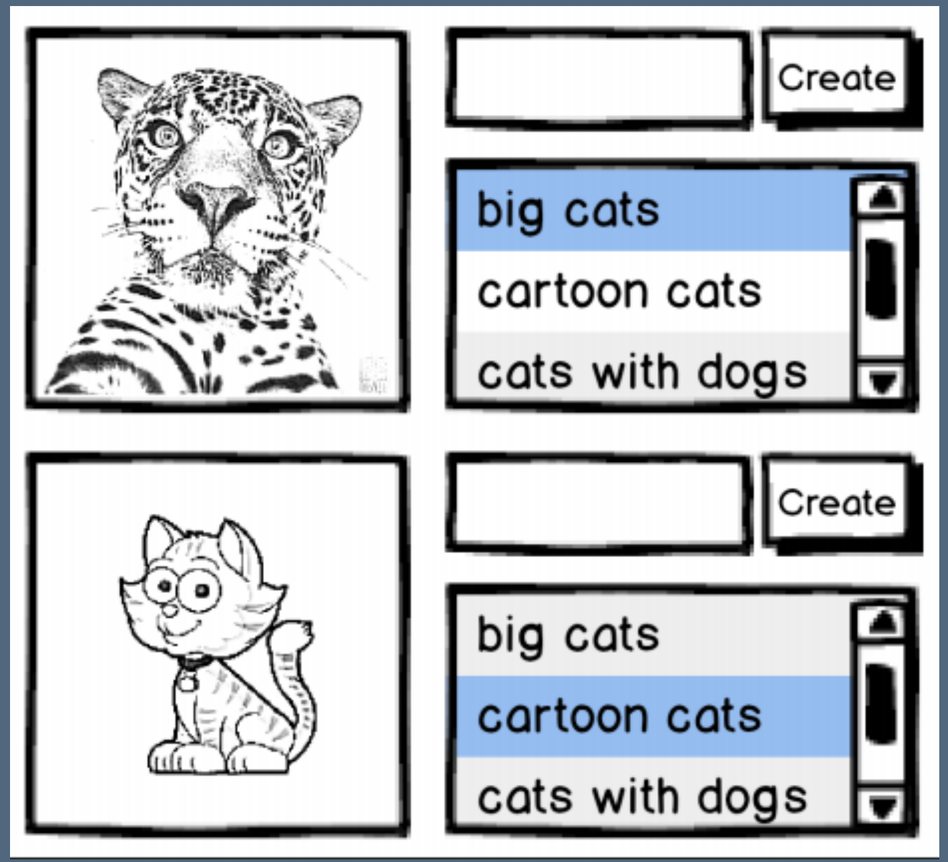


Revising your training as you go

[Chang, Amershi and Kamar 2017]

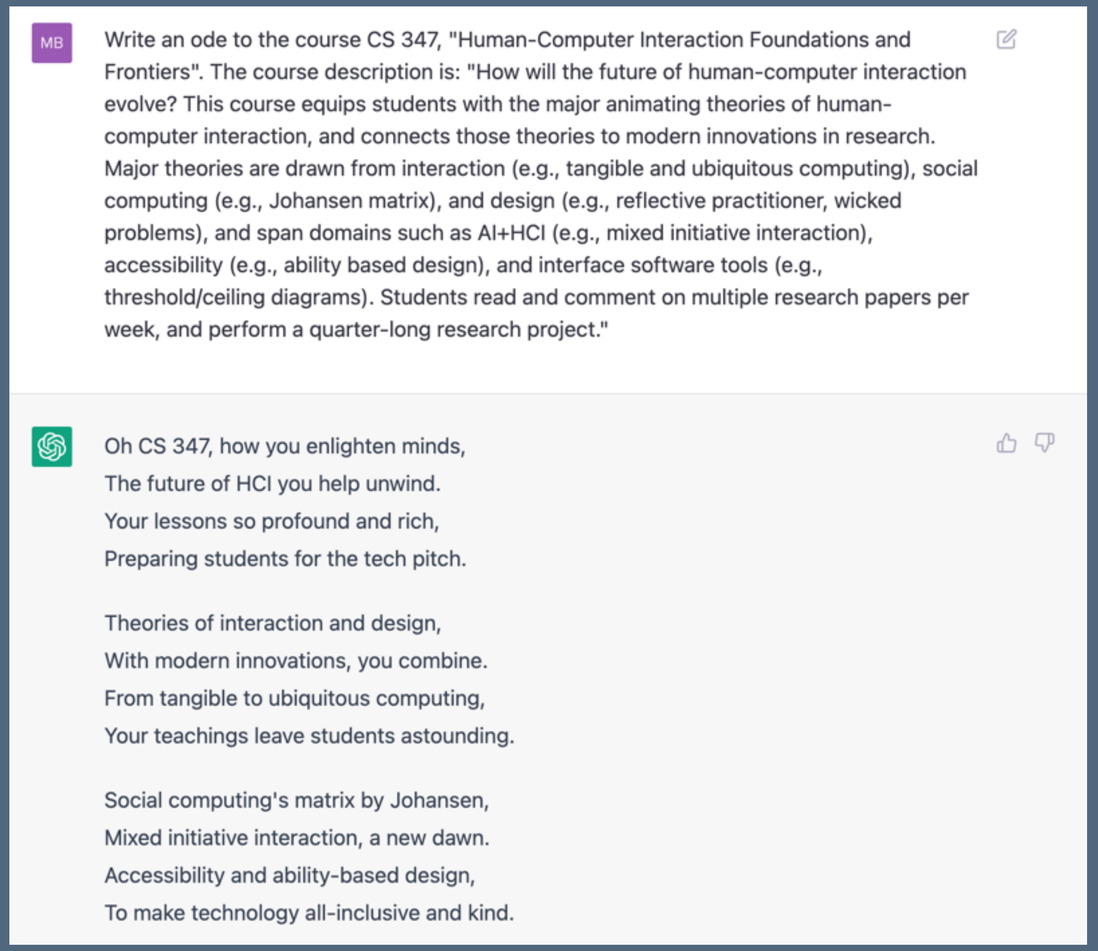
Facilitate concept evolution through a "could be" category that allows clustering into subcategories you can change labels for





More recently: prompting

In-context learning allows end users to write what they want:

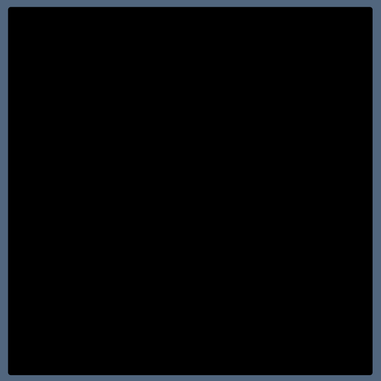


Control remains an open problem

If I can't figure out how to cross the gulf of execution through the prompt, how do I convey my intent?

Unpredictable black boxes are terrible user interfaces

Picture of a Professor named Maneesh Agrawala

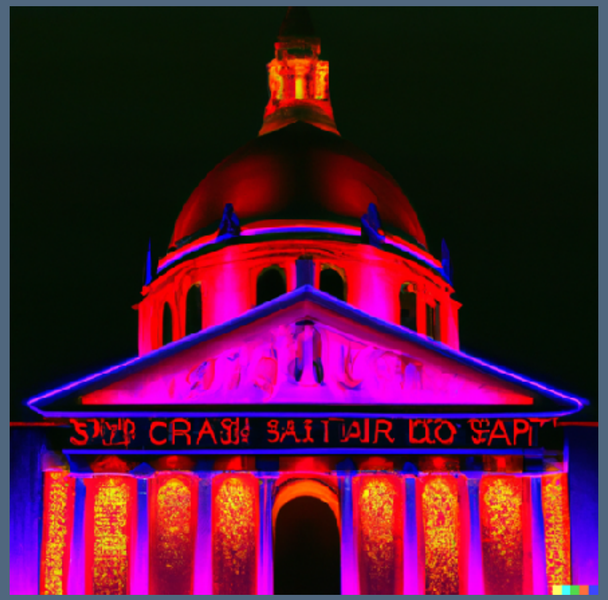


DALL-E2


カクテル
基礎の上に
お手持ちの鳥口
おいしい料理
人造石研磨
Wine


academy 3F
歌舞伎市場
八方
寿司
ホルモン
はなけん
YAKINIKU Bar
NO MEAT, NO LIFE.
2nd
1F
3F
菓子屋
駄菓子
居酒家日
DAGASHI & SAKAYA
Ennichi
NO MEAT, NO LIFE. 3rd
03-3232-4888
NO MEAT, NO LIFE. 2nd
03-3209-1129
相席無料
すし
STAFF

stanford memorial church with neon signage in the style of bladerunner



Iteration 1

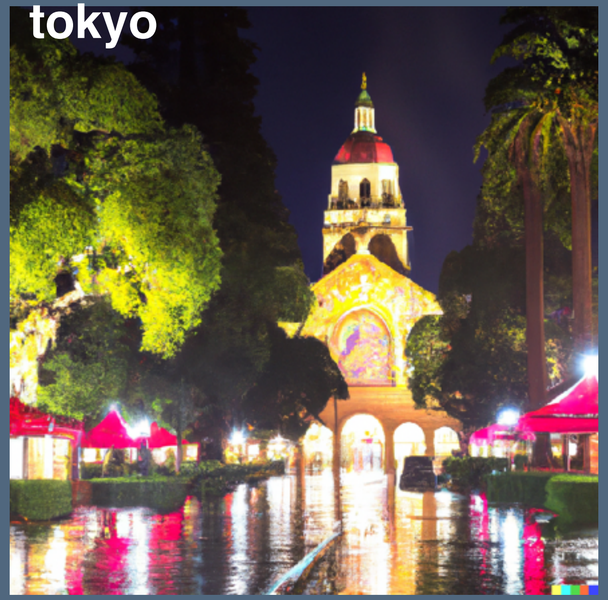
stanford memorial church **and main quad with palm trees** in the style of bladerunner

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

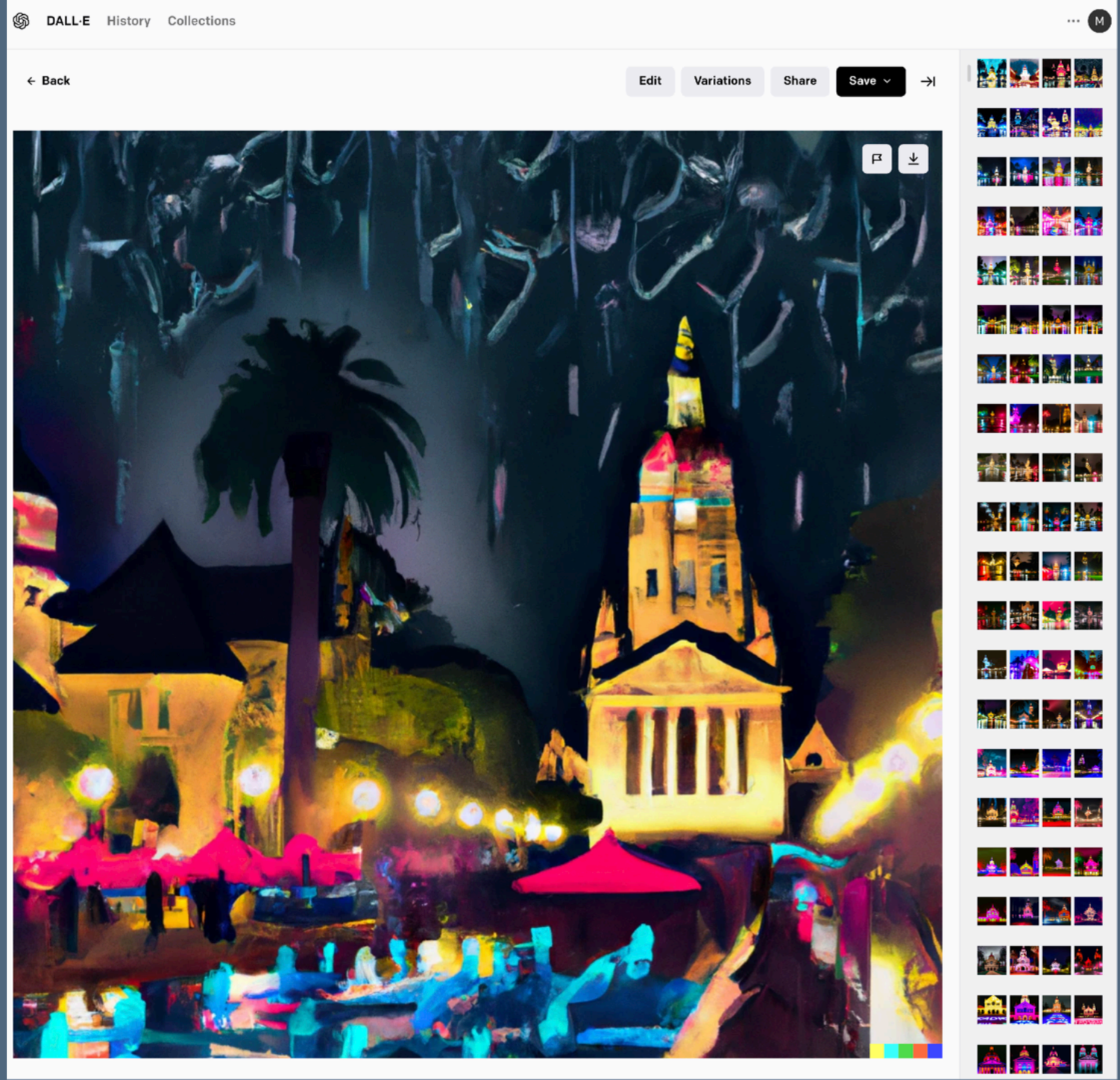
nighttime rain stanford memorial church and main quad with palm trees, night market food stalls and neon signs **like downtown tokyo**



Iteration 17

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**

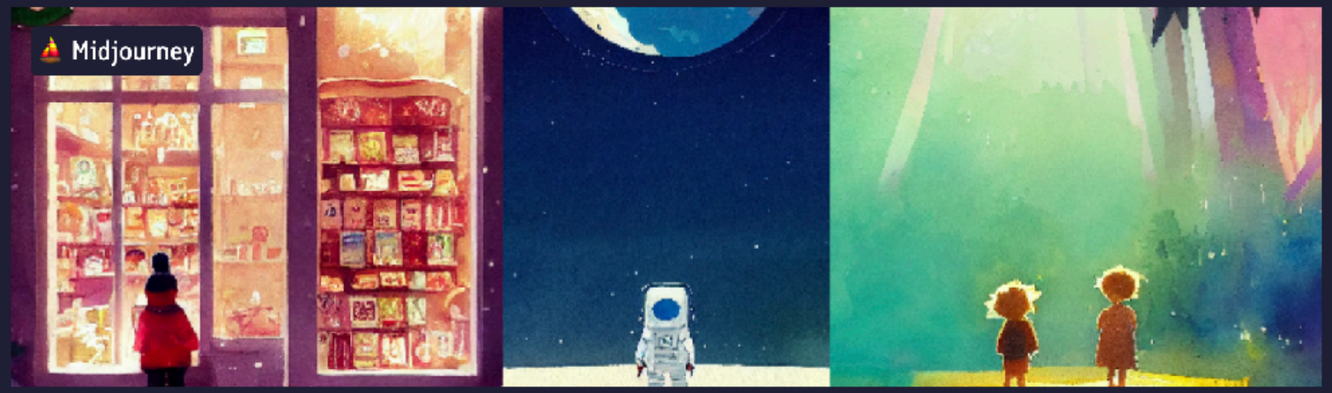
Iteration 21



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



Beautiful Watercolor Illustrations

505 @imagineer

Generates beautiful watercolor illustrations with undefined figures, in a consistent style. Ideal for illustrating stories, tales or blogs with vivid and colorful watercolor images. You can select the proportions of each generated illustration.

$1.99

Get Prompt

After purchasing, you will gain access to the prompt file, which you can use with Midjourney. You must already have access to Midjourney to use this prompt.

Why Johnny Can't Prompt

[Zamfirescu-Pereira et al. 2023]

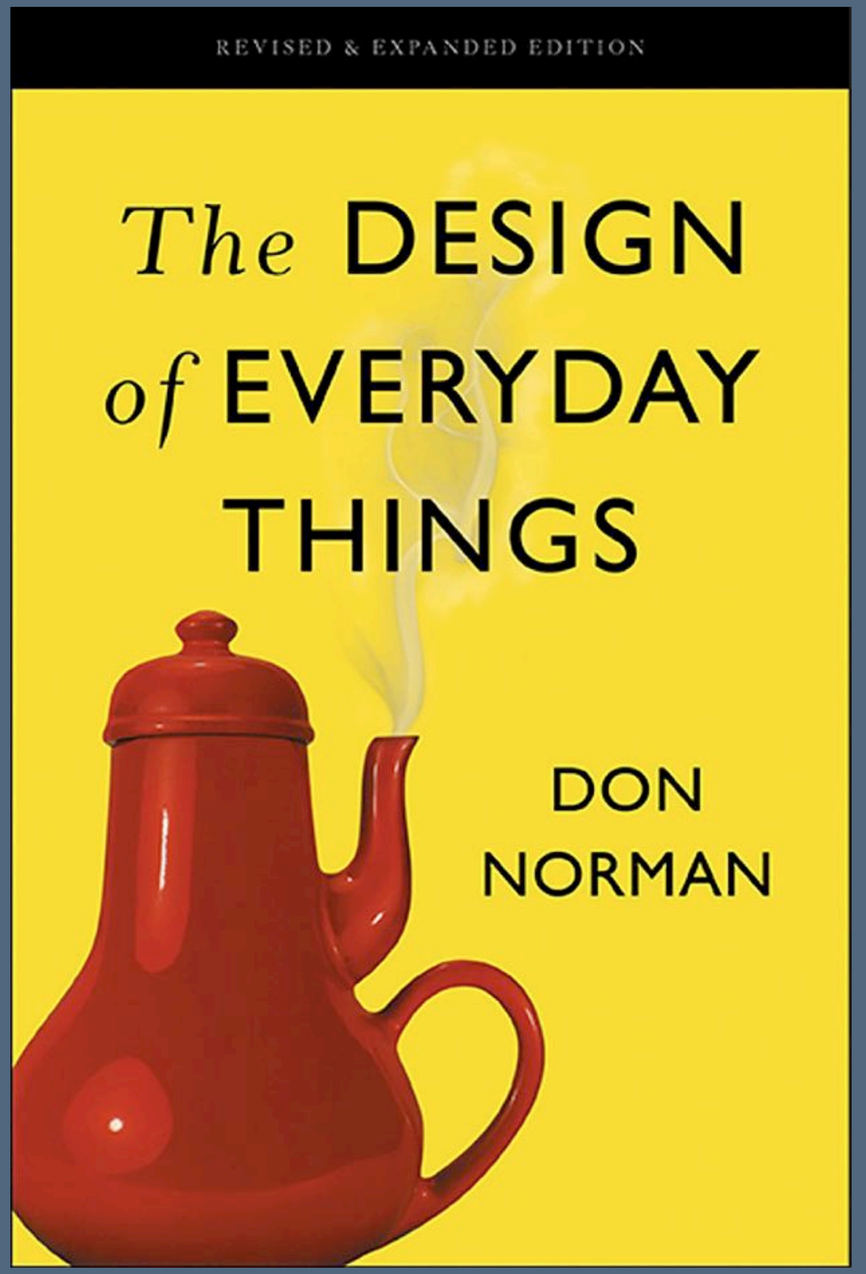
YOU READ THIS

Prompters **don't know what AI can/cannot do**. So need examples or instructions on how to proceed. Consistent with [Yang 2020].

Prompters **over-generalize** from a few examples, or errors (give up early).

Prompters **anthropomorphize** and filter expectations based on human-human interactions.

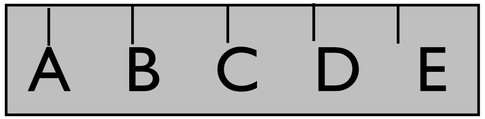
> Gave direct instructions instead of providing in-context examples. Even when instructed by human researcher to give examples.
>
> Some prompters expected AI to understand instructions the way a human would (e.g. instruction: 'do not use ABC', result: AI uses ABC verbatim in response)

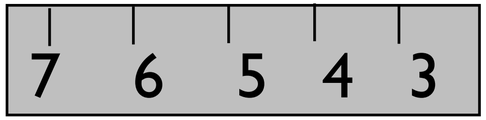

REVISED & EXPANDED EDITION
The DESIGN of EVERYDAY THINGS
DON NORMAN

freezer

fresh food

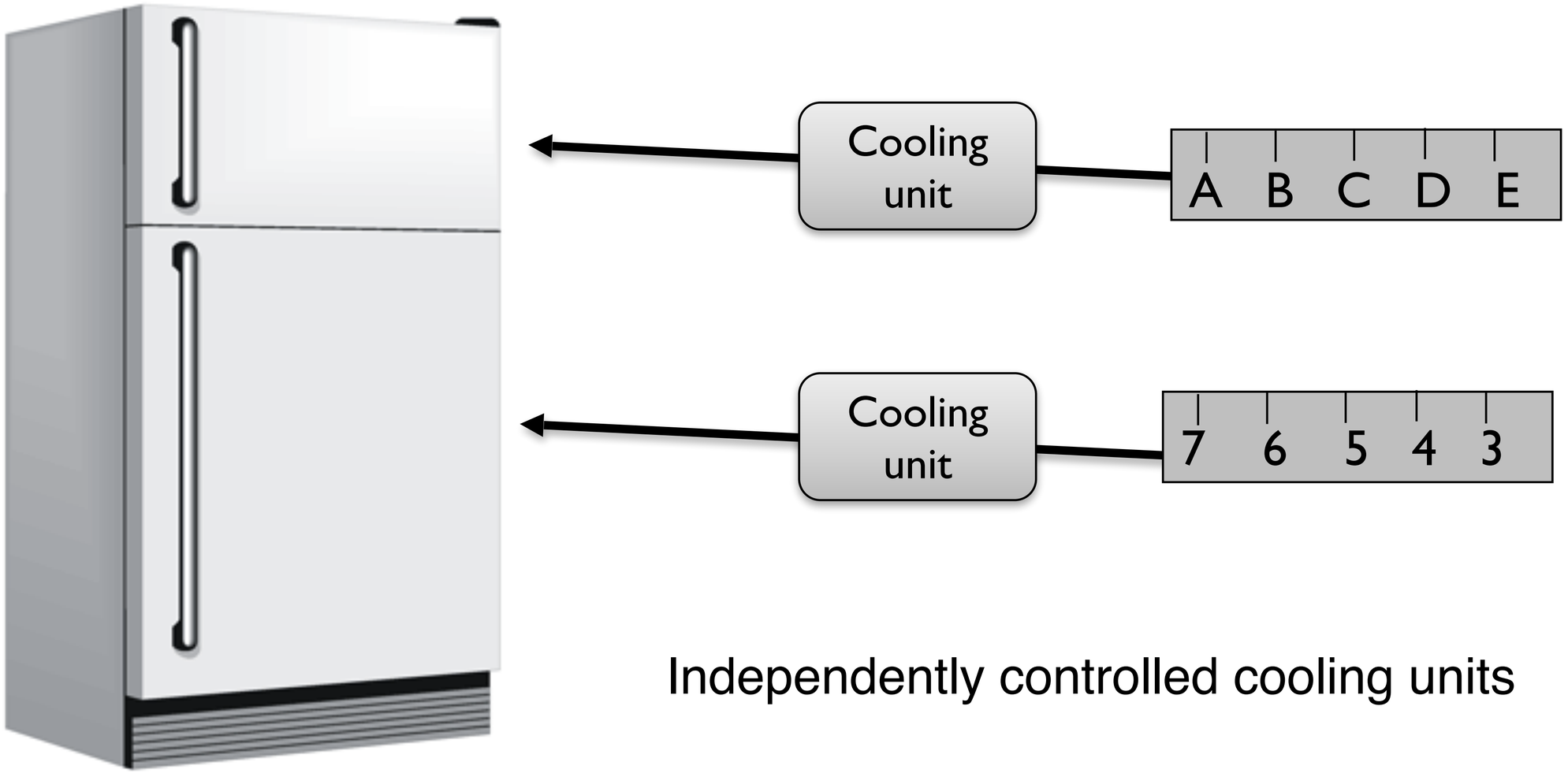
Normal Settings	C and 4
Colder Fresh Food	C and 5-6
Coldest Fresh Food	B and 7
Colder Freezer	D and 6-7
Warmer Fresh Food	C and 3-1
OFF (both)	



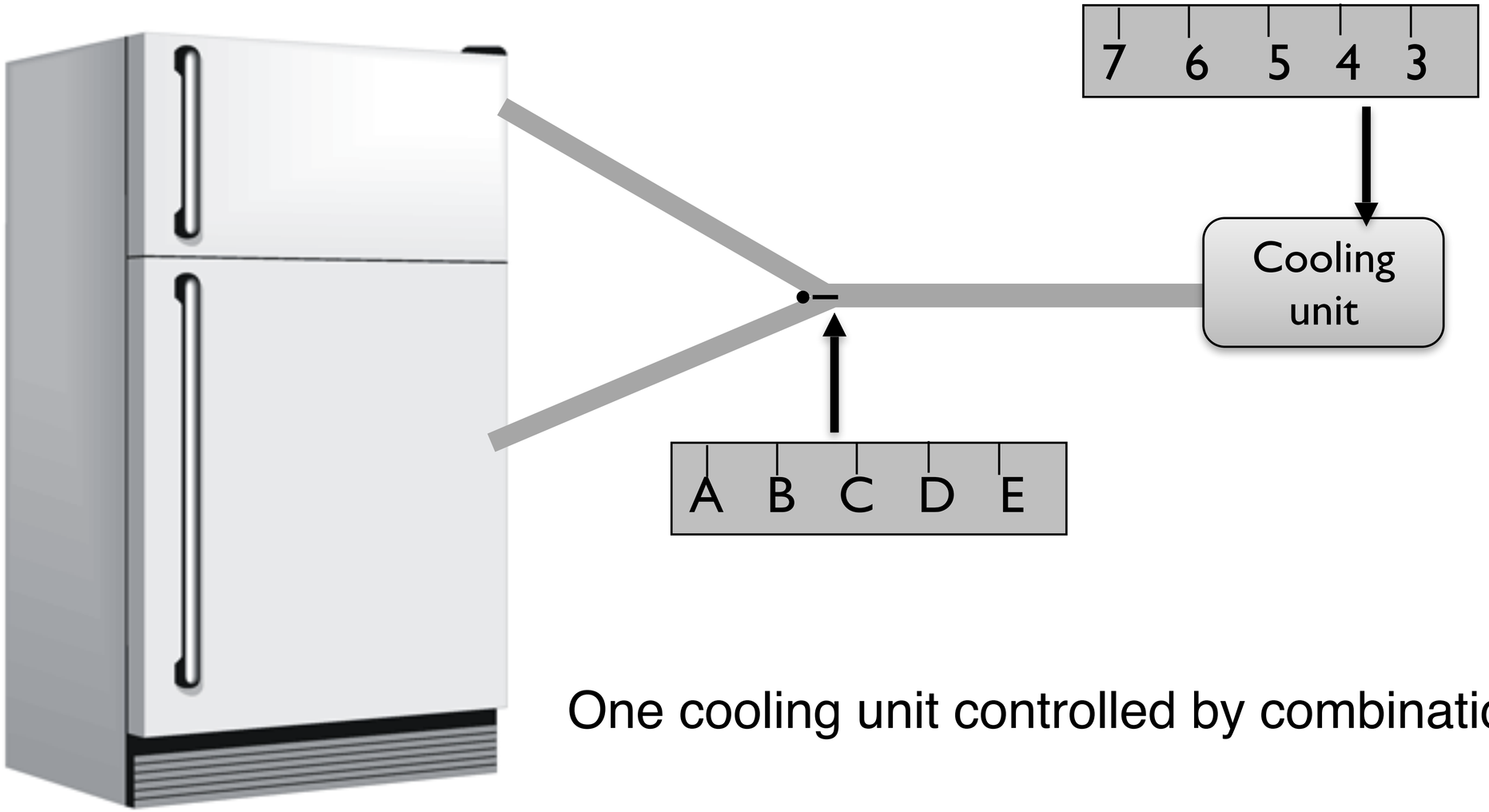


Freezer

Fresh Food



Independently controlled cooling units



One cooling unit controlled by combination of inputs

A **good conceptual model** let's users **predict** how **input controls** affect the **output**

When the **conceptual model** is **not predictive,** users resort to **trial-and-error**

It is ***our* job** as AI tool builders to provide interfaces that **let users build predictive conceptual models**

Picture of a cool, young Computer Science Professor named Maneesh Agrawala

DALL-E2

AI black boxes are terrible interfaces

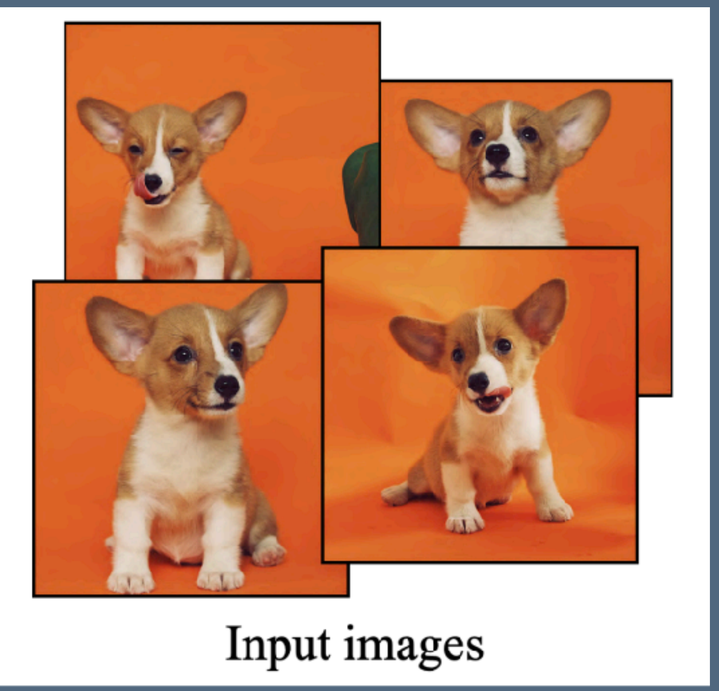
- Does “cool” imply a sportscoat?
- Does “picture” generate a photograph?
- Cannot predict how input prompt affects output image

Conversational Interactions with ChatGPT

Support for turn-taking and context enables some common ground

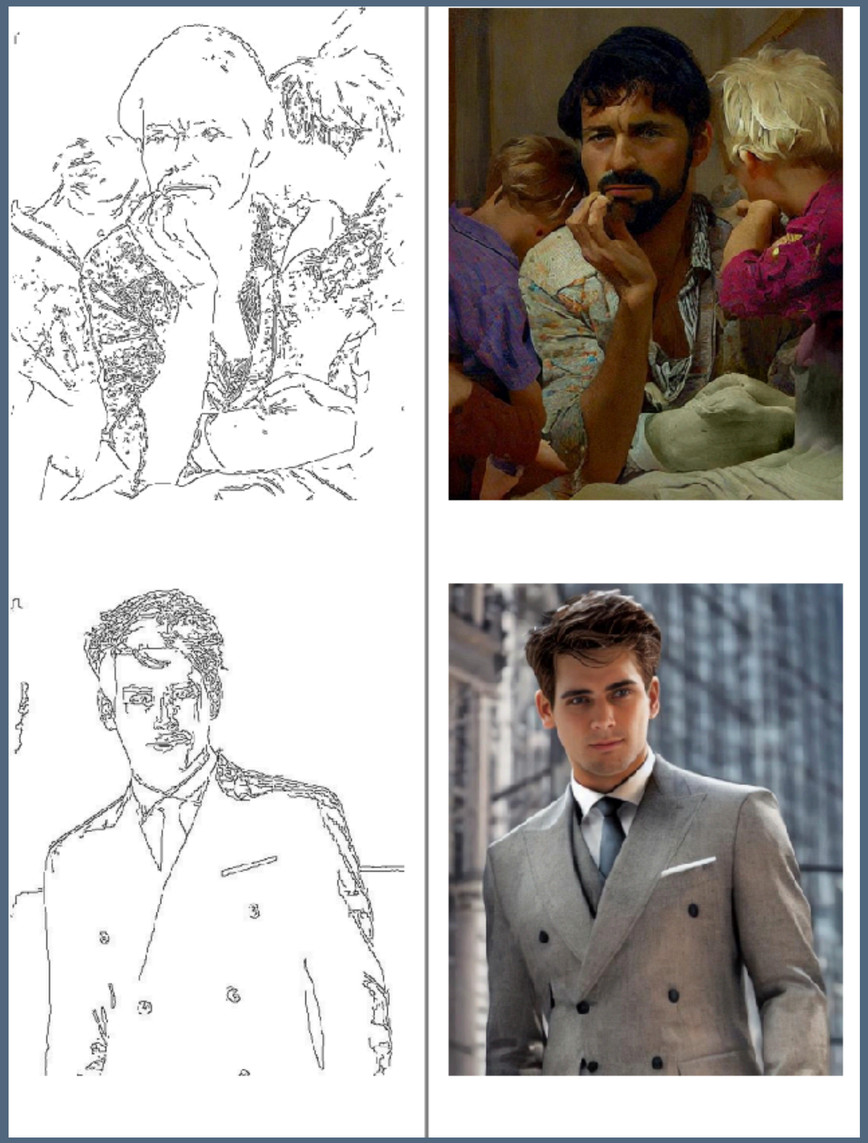
- AI and human **can refer to concepts** from earlier exchanges **to refine them**
- But **refinement is one-sided**. AI doesn't ask for refinement – human adds it
- **AI model does not immediately learn** (or update weights) from the conversation.
- **Grounding seems shallow** (unclear what ChatGPT knows and doesn't know/remember)

Establishing Common Ground



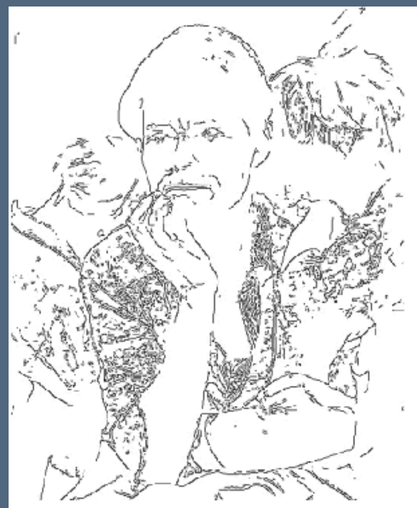
DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation [Ruiz 2022]

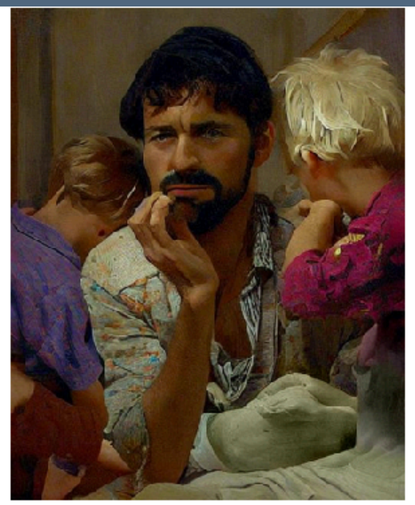
Adding Conditional Control to Text-to-Image Diffusion Models [Zhang 2023]

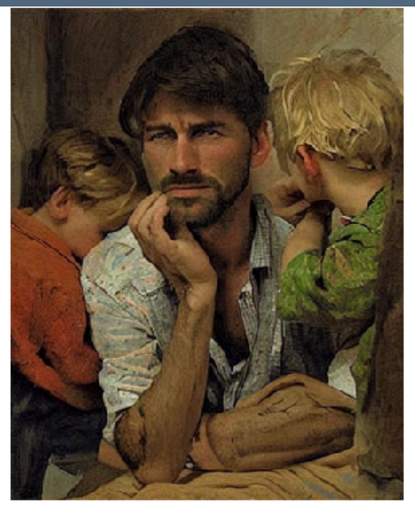


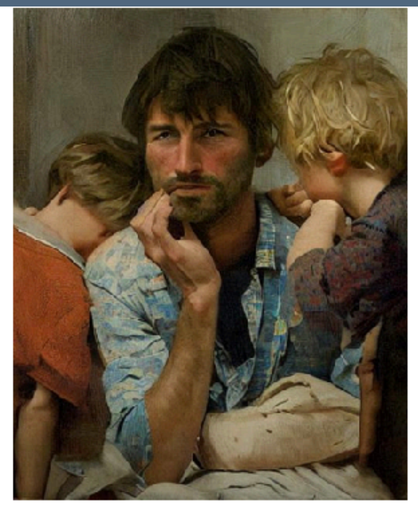
Adding Conditional Control to Text-to-Image Diffusion Models [Zhang 2023]

Adding Conditional Control to Text-to-Image Diffusion Models [Zhang 2023]

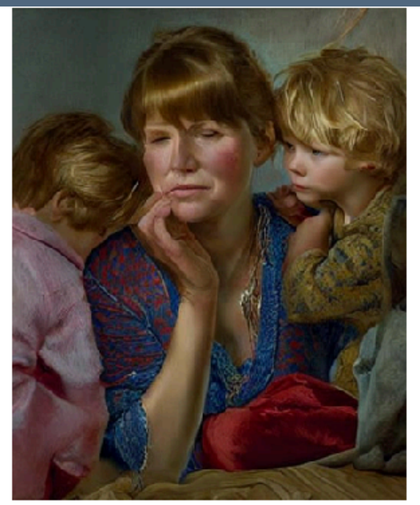




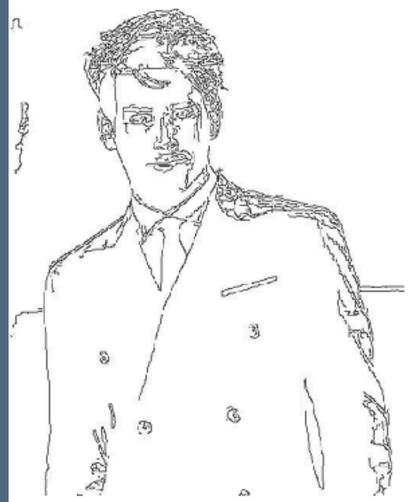




"a man with beard sitting with two children"



"mother and two boys in a room, masterpiece, artwork"



"a man in a suit and tie"

"a man in a white suit and tie"

Adding Conditional Control to Text-to-Image Diffusion Models [Zhang 2023]

Dealing with Ambiguity of Spatial Language

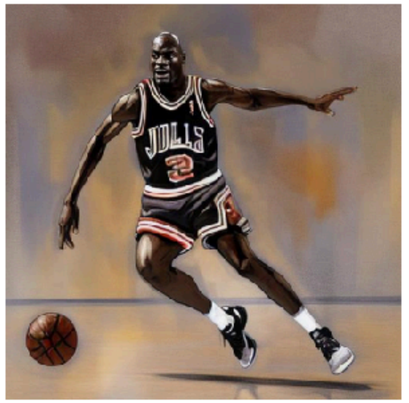
Idea: User provides *conditioning* image that puts spatially localized constraints on the output image

Adding Conditional Control to Text-to-Image Diffusion Models [Zhang 2023]

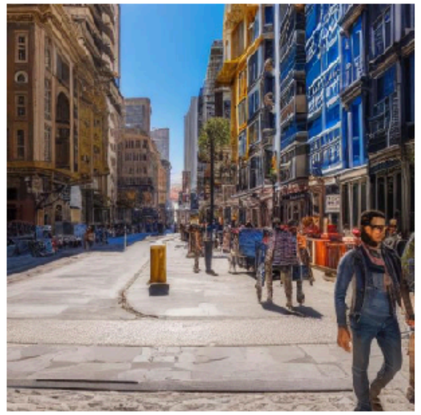
Dealing with Ambiguity of Spatial Language



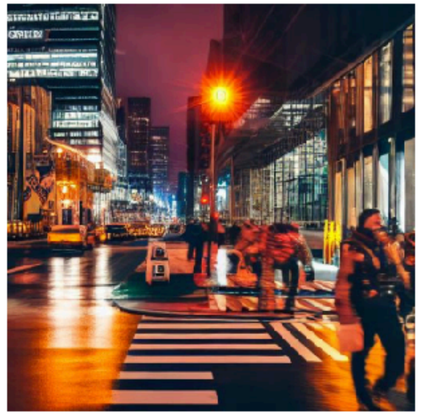

"a woman dancing near a street corner"




"artwork of Michael Jordan playing basketball"





Idea: User provides *conditioning* image that puts spatially localized constraints on the output image

Adding Conditional Control to Text-to-Image Diffusion Models [Zhang 2023]

Summary

Intelligence augmentation aims to place AI in context by using it to amplify our own abilities

Debates rage about the levels of autonomy to grant to AIs: from fully autonomous **agents** that act on the person's behalf, to **direct manipulation** that always leaves the user in full control

Mixed initiative interaction splits the difference by asking, acting, or doing nothing based on its confidence and utility

When users cannot predict how input controls affect outputs the interface, the results can be frustrating and terrible

References

Adar, Eytan. “Bounced checks at the UI/AI intersection.” Stanford HCI Seminar. 2018. https://www.youtube.com/watch?v=11UKXaELg8M

Breazeal, Cynthia. Designing sociable robots. MIT press, 2004.

Chang, Joseph Chee, Saleema Amershi, and Ece Kamar. "Revolt: Collaborative crowdsourcing for labeling machine learning datasets." Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017.

Christin, Angèle. "Algorithms in practice: Comparing web journalism and criminal justice." Big data & society 4.2 (2017): 2053951717718855.

Dragan, Anca D., Kenton CT Lee, and Siddhartha S. Srinivasa. "Legibility and predictability of robot motion." 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2013.

Engelbart, Douglas C. "Augmenting human intellect: A conceptual framework." Menlo Park, CA 21 (1962).

Fails, Jerry Alan, and Dan R. Olsen Jr. "Interactive machine learning." Proceedings of the 8th international conference on Intelligent user interfaces. 2003.

Fogarty, James, et al. "CueFlik: interactive concept learning in image search." Proceedings of the sigchi conference on human factors in computing systems. 2008.

Heer, Jeffrey. "Agency plus automation: Designing artificial intelligence into interactive systems." Proceedings of the National Academy of Sciences 116.6 (2019): 1844-1850.

References

Horvitz, Eric. "Principles of mixed-initiative user interfaces." Proceedings of the SIGCHI conference on Human Factors in Computing Systems. 1999.

Ko, Hyung-Kwon, et al. "We-toon: A Communication Support System between Writers and Artists in Collaborative Webtoon Sketch Revision." Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. 2022.

Krishna, Ranjay, et al. "Socially situated artificial intelligence enables learning from human interaction." Proceedings of the National Academy of Sciences 119.39 (2022): e2115730119.

Maes, Pattie. "Agents that reduce work and information overload." Readings in human–computer interaction. Morgan Kaufmann, 1995. 811-821.

Mok, Brian, et al. "Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles." 2015 IEEE 18th international conference on intelligent transportation systems. IEEE, 2015.

Shneiderman, Ben, and Pattie Maes. "Direct manipulation vs. interface agents." interactions 4.6 (1997): 42-61.

Shneiderman, Ben. Human-centered AI. Oxford University Press, 2022.

Vaccaro, Michelle, Abdullah Almaatouq, and Thomas Malone. "When combinations of humans and AI are useful: A systematic review and meta-analysis." Nature Human Behaviour (2024): 1-11.

Yang, Qian, et al. "Re-examining whether, why, and how human-AI interaction is uniquely difficult to design." Proceedings of the 2020 chi conference on human factors in computing systems. 2020.