Human-Centered Artificial Intelligence

CS 347

Maneesh Agrawala

Last time

Collaboration is hard: physical **distance matters**.

Tools can try to mitigate the effects of distance, but we are limited by the **socio-technical gap**.

Crowdsourcing gives up on tight teamwork in favor of structured contributions through open call and at massive scale

✓ Social Computing

Unit 3

social media
collaboration

Where we go from here

so far	Ubiquitous Computing, Design, Social Computing
week 5	Human-Centered AI
week 6	Cognition/Visualization
week 7	Software Tools/Content Creation
week 8	Critical Theory/Simulating People
week 9	Methodology
week 10	History

Human Centered AI

Unit 4

human-centered AI
working with unpredictable black boxes

Today

AI vs. IA

Direct manipulation vs. Agents

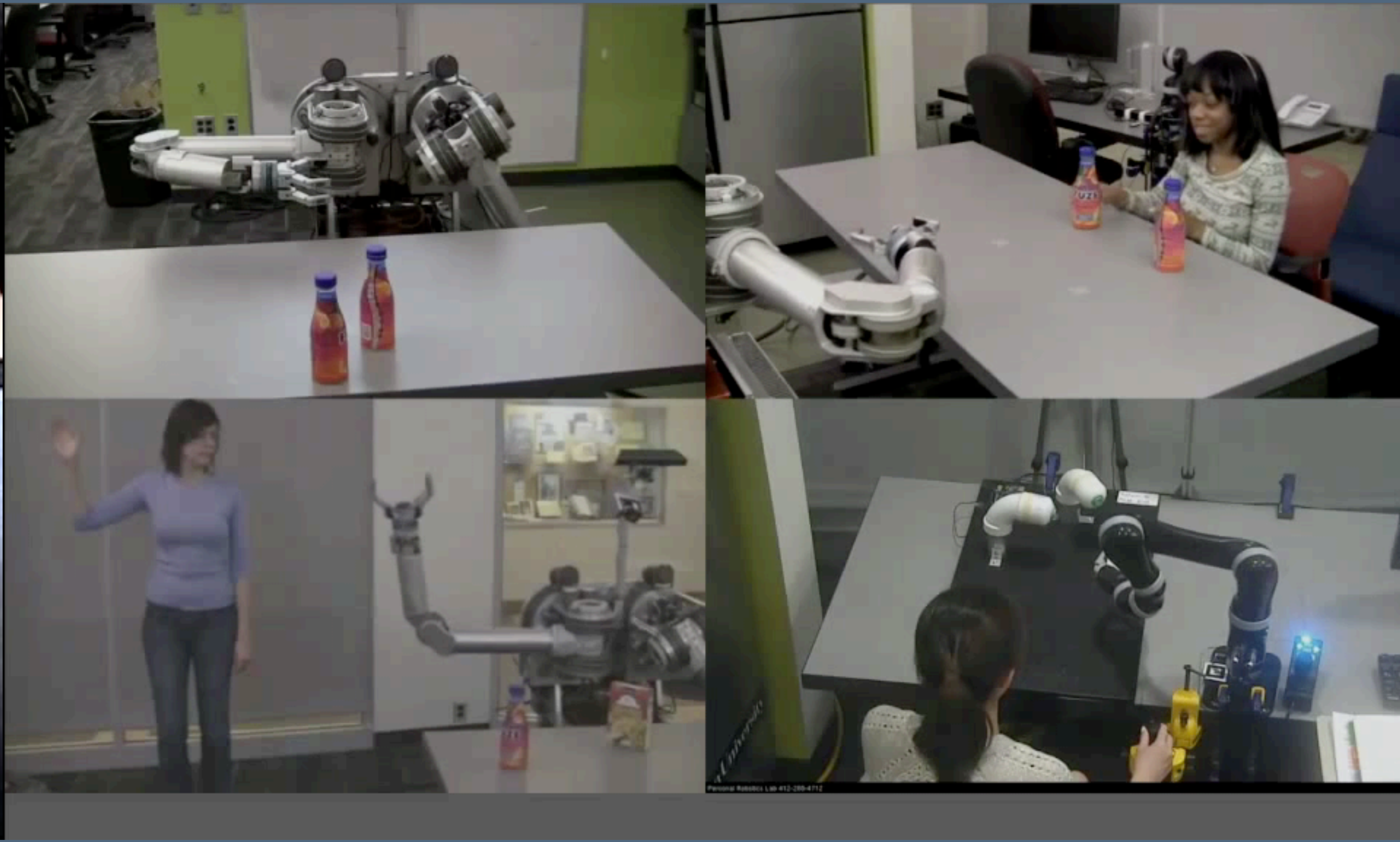
Mixed-initiative interaction

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]

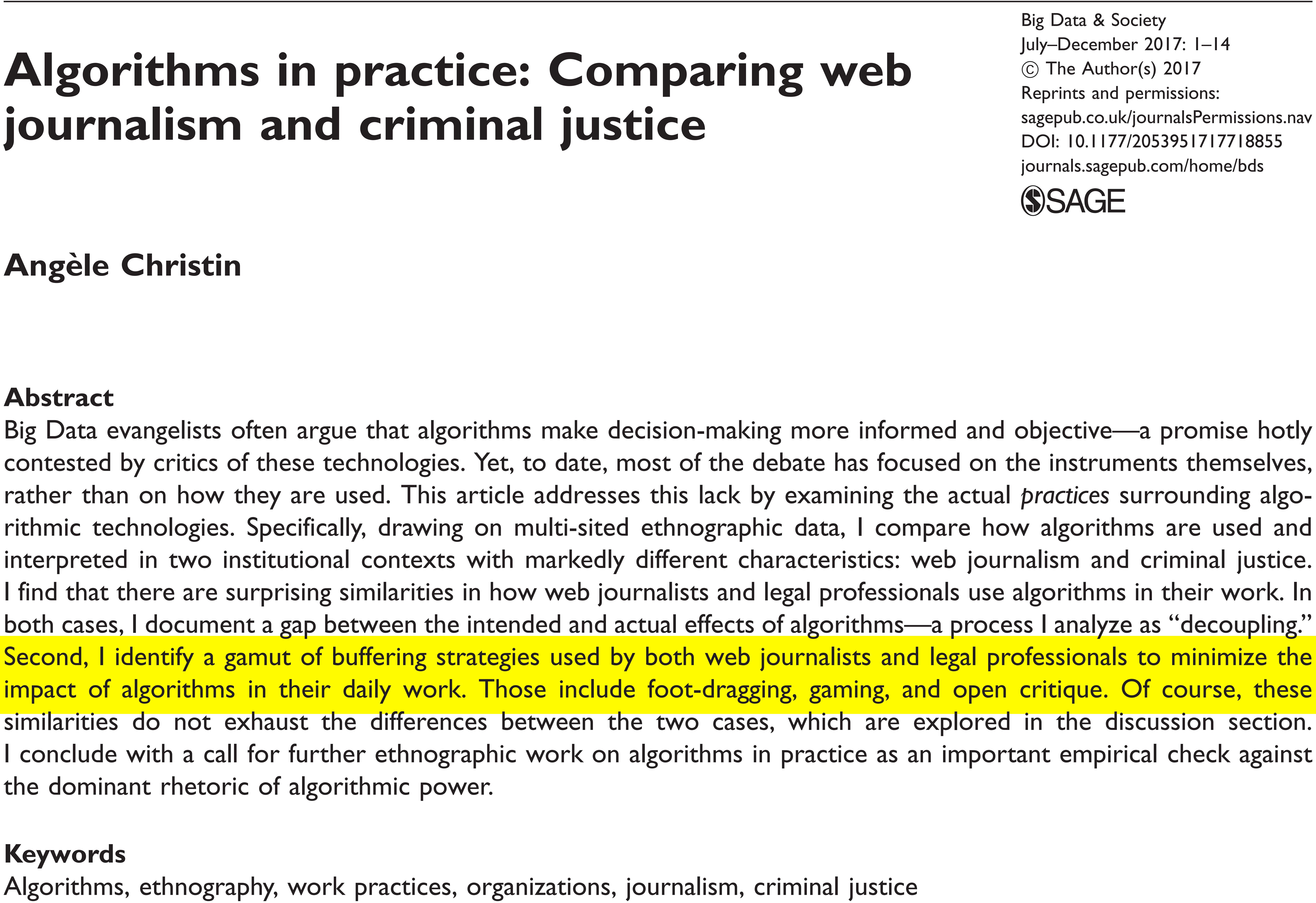
Intelligence Augmentation

AI vs. IA

A reaction to:

“AI will replace human intelligence”

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



Big Data & Society
July–December 2017: 1–14
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2053951717718855
journals.sagepub.com/home/bds
SAGE

Algorithms in practice: Comparing web journalism and criminal justice

Angèle Christin

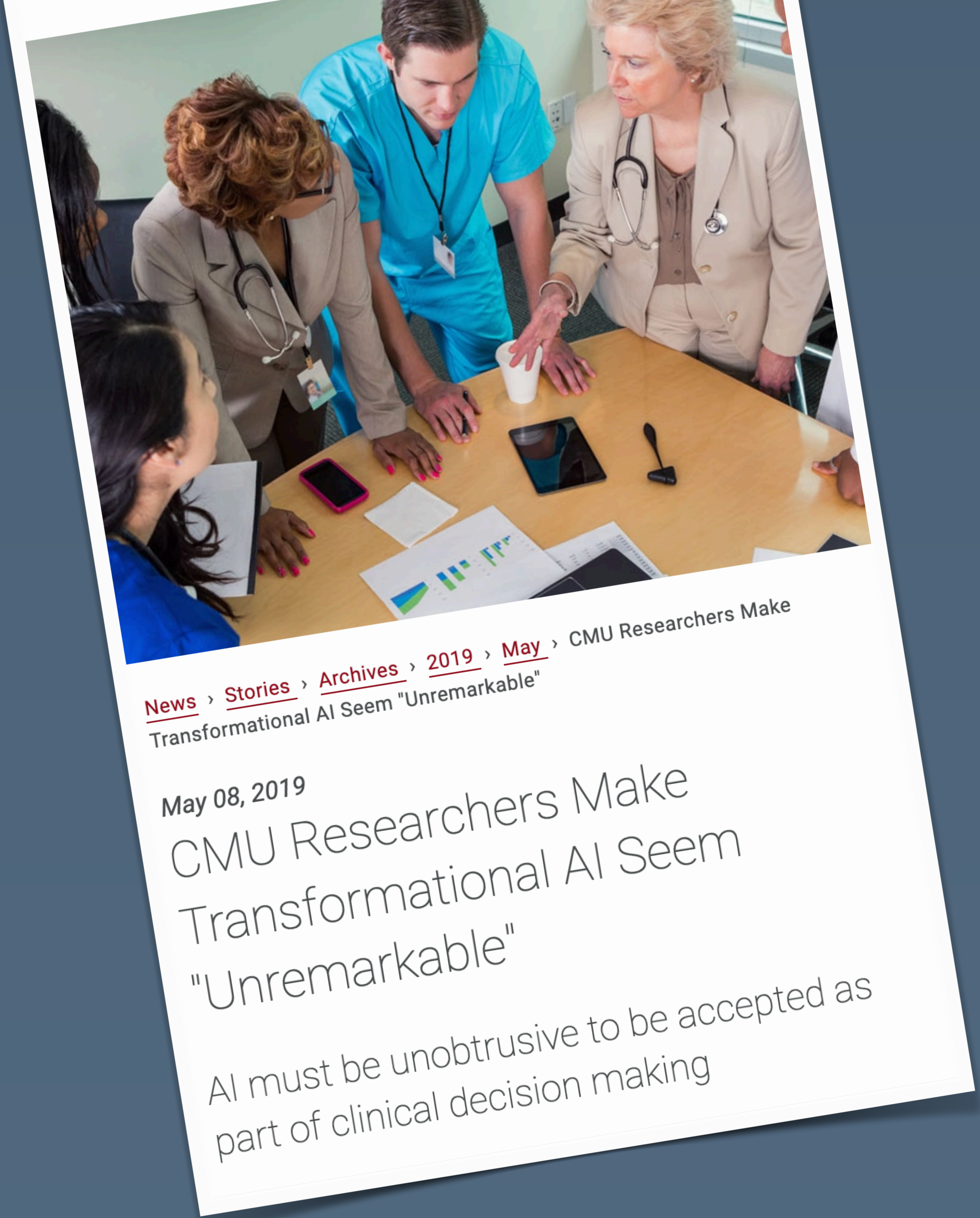
Abstract

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.

Keywords

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

If you try thoughtlessly…

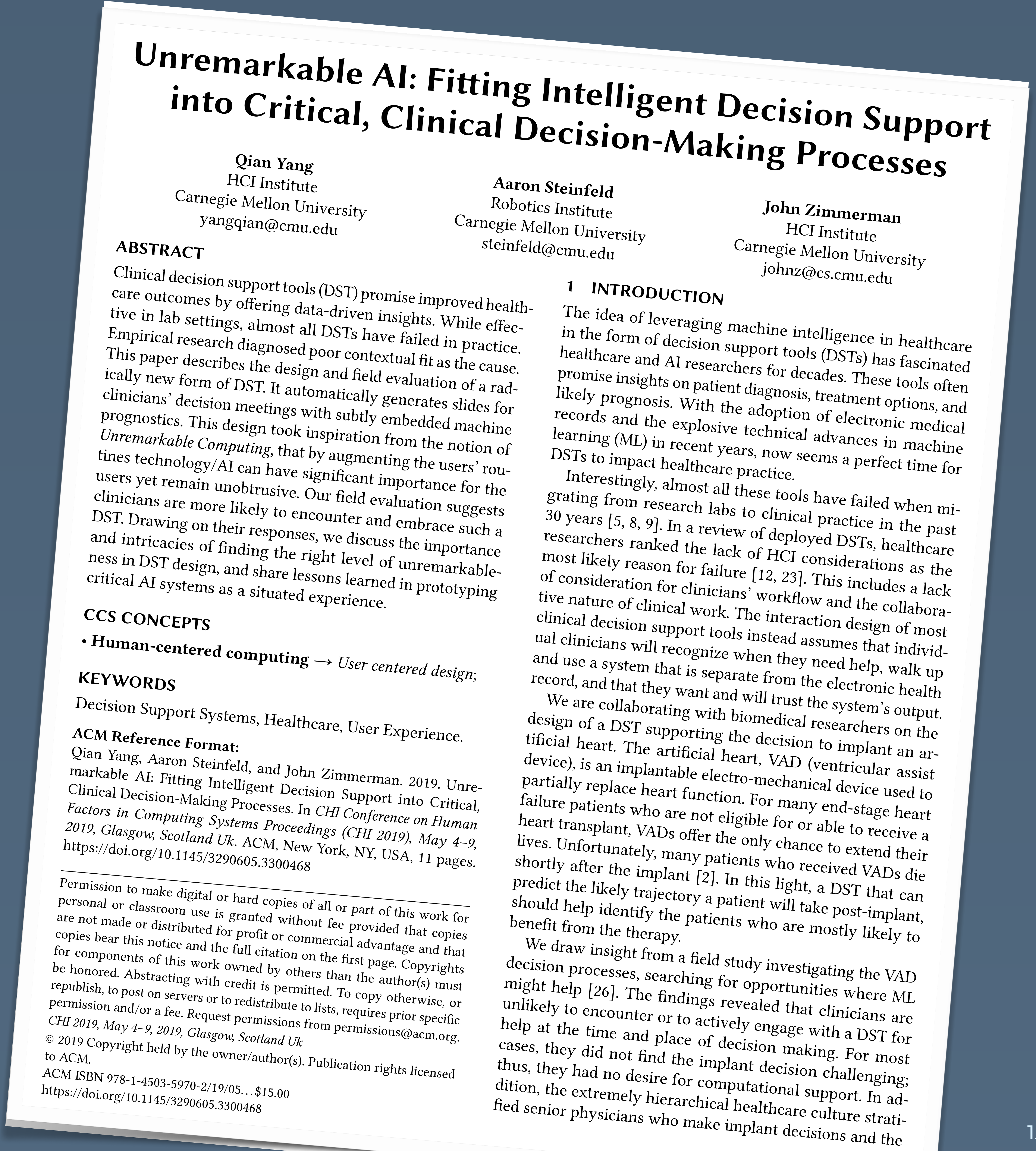


News › Stories › Archives › 2019 › May › CMU Researchers Make Transformational AI Seem "Unremarkable"

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

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

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300468

Our trust isn't calibrated

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

What if the algorithm just suggests the answer to you?

We often get influenced by the AI's suggestion and rely on it when we shouldn't [Buçinca, Malaya, and Gajos 2021]

But surely if the algorithm explains its reasoning?

Doesn't help, unless the explanation takes almost no effort to verify [Vasconcelos et al. 2023]

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*

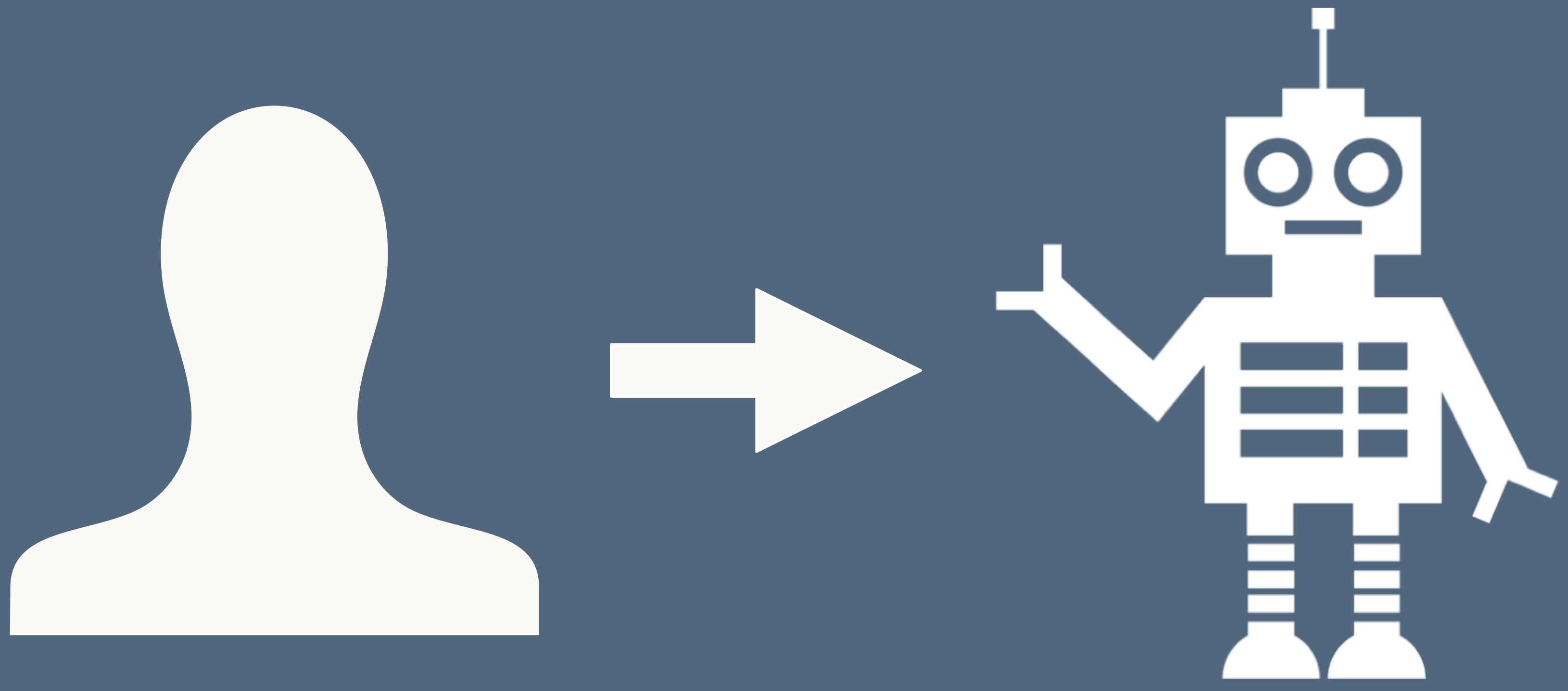
STANFORD RESEARCH INSTITUTE

MENLO PARK, CALIFORNIA

*SRI

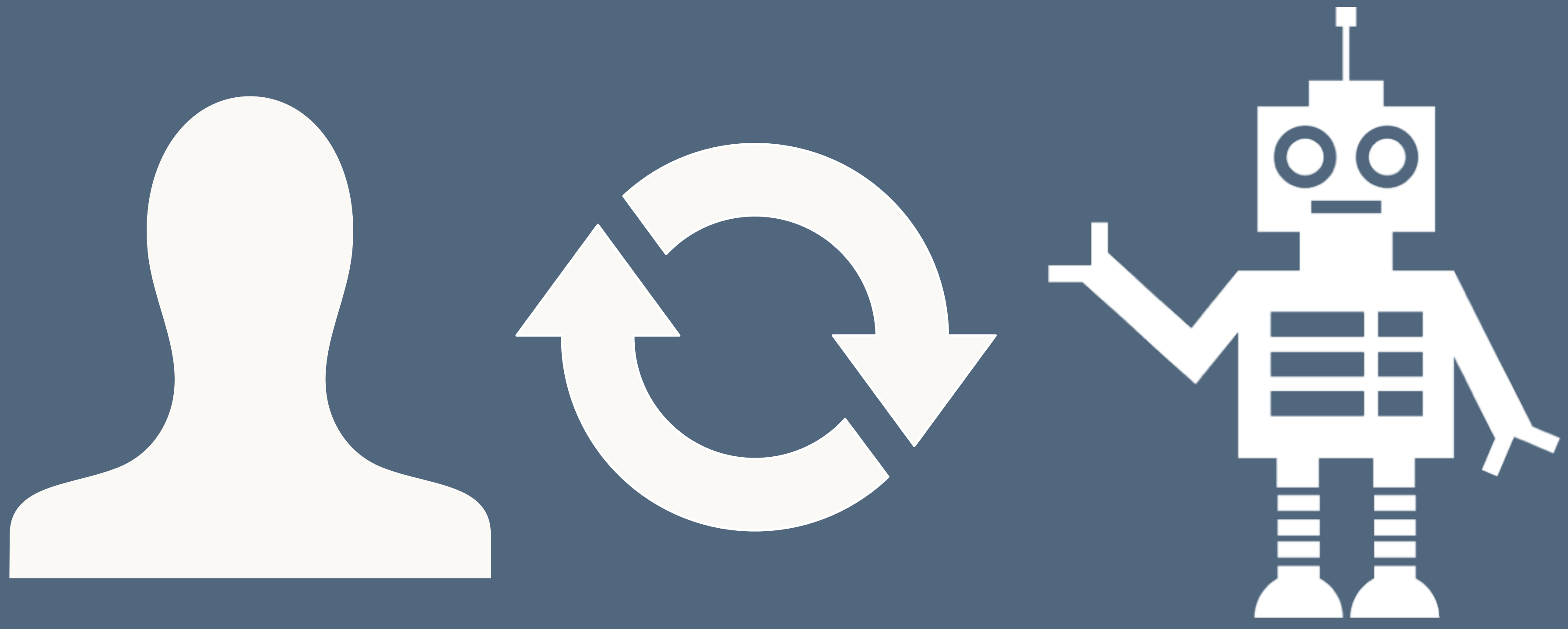
SRI

Artificial Intelligence



Replace human intelligence with artificial intelligence

Intelligence Augmentation



Augment human intelligence with artificial intelligence

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

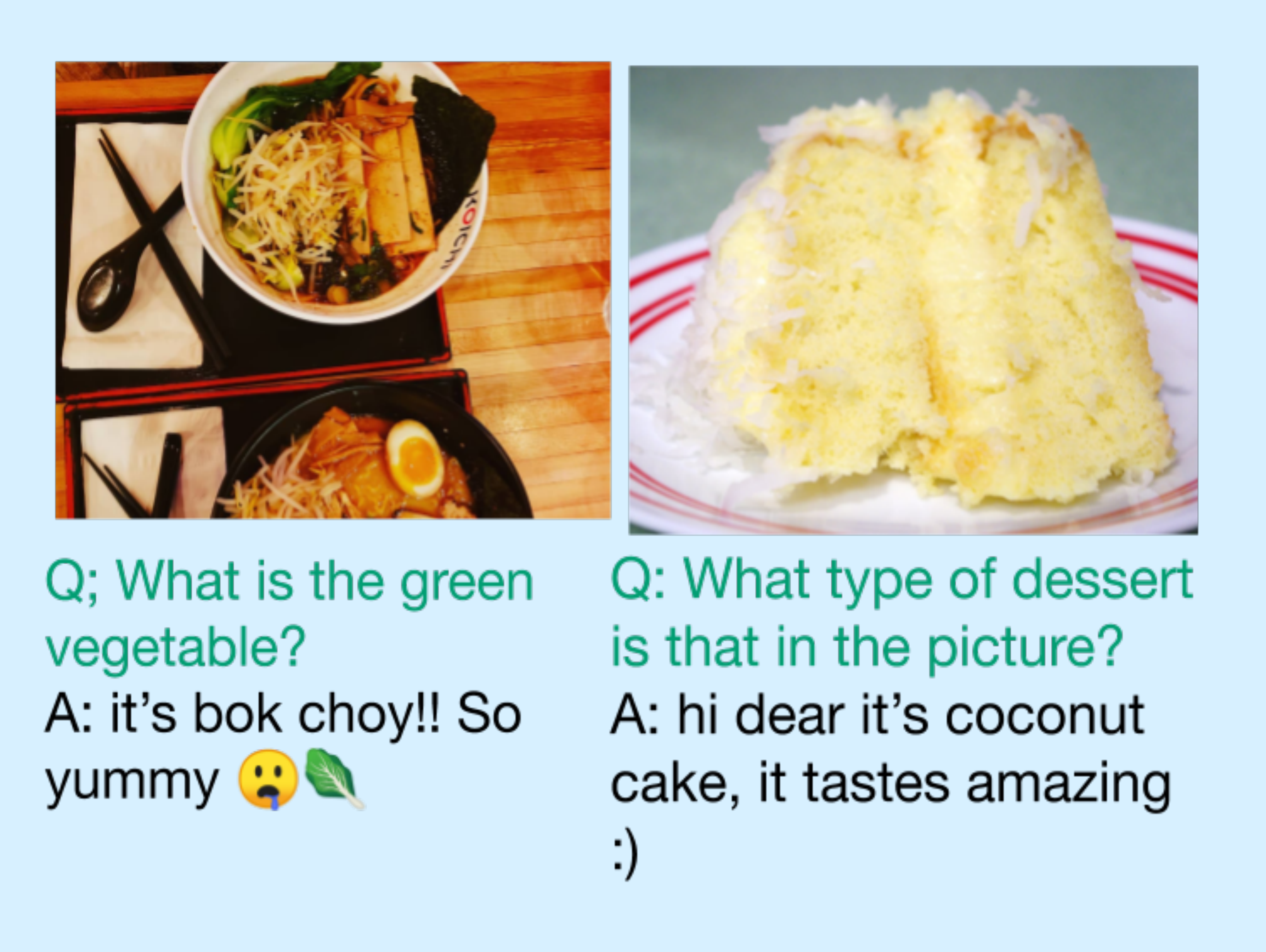
Users should always have full control, even as automation increases

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]



Q; What is the green vegetable?
A: it's bok choy!! So yummy 🤤🥬

Q: What type of dessert is that in the picture?
A: hi dear it's coconut cake, it tastes amazing :)

Direct manipulation

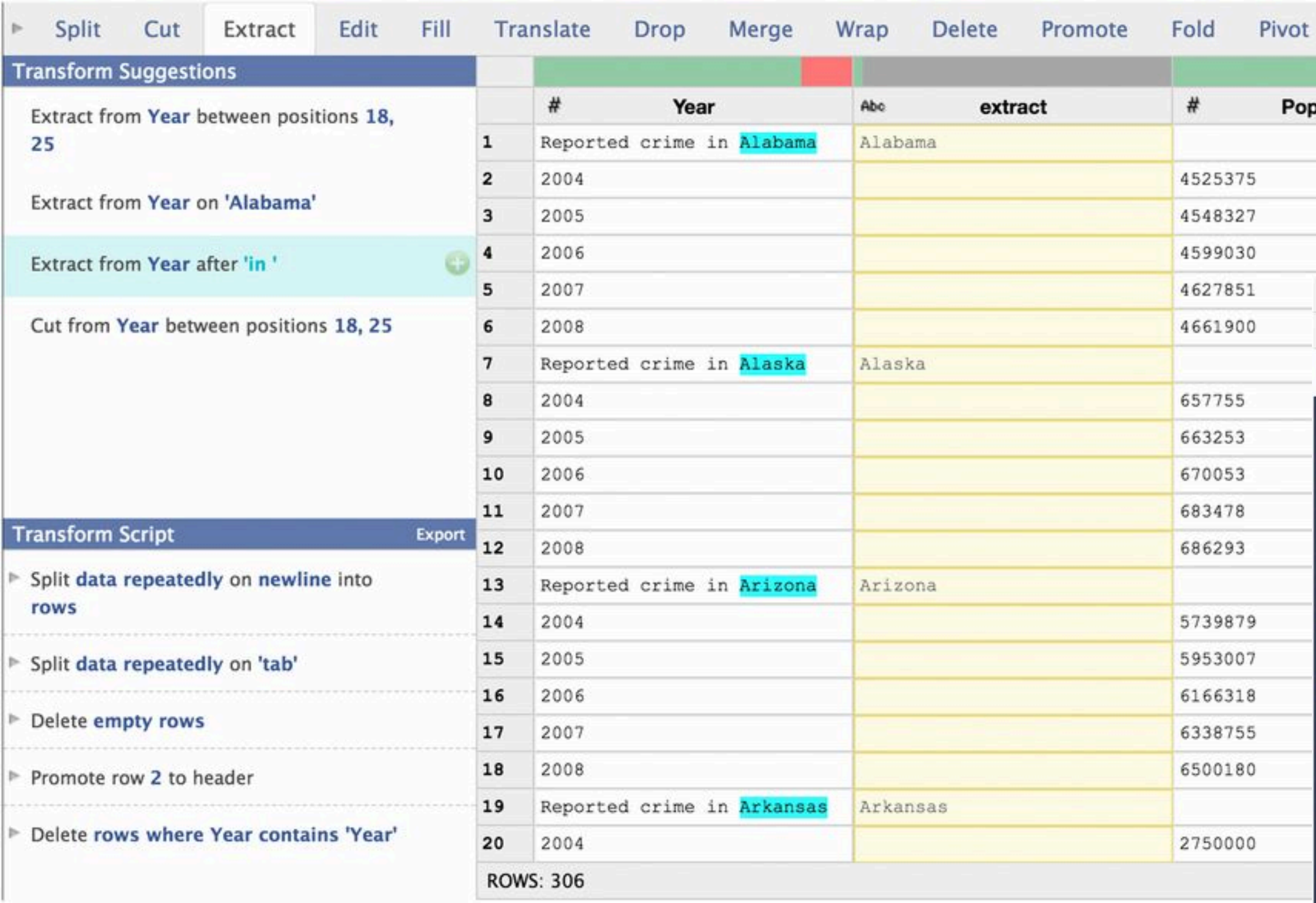
Shneiderman: it is possible to maintain high levels of user control even as automation increases [Shneiderman 2022]

Control

	Low	High
High	bicycle piano	camera
Low	music box	airbag pacemaker

Automation

Agency plus automation [Heer 2019]



Transform Suggestions

- Extract from Year between positions 18, 25
- Extract from Year on 'Alabama'
- Extract from Year after 'in '
- Cut from Year between positions 18, 25

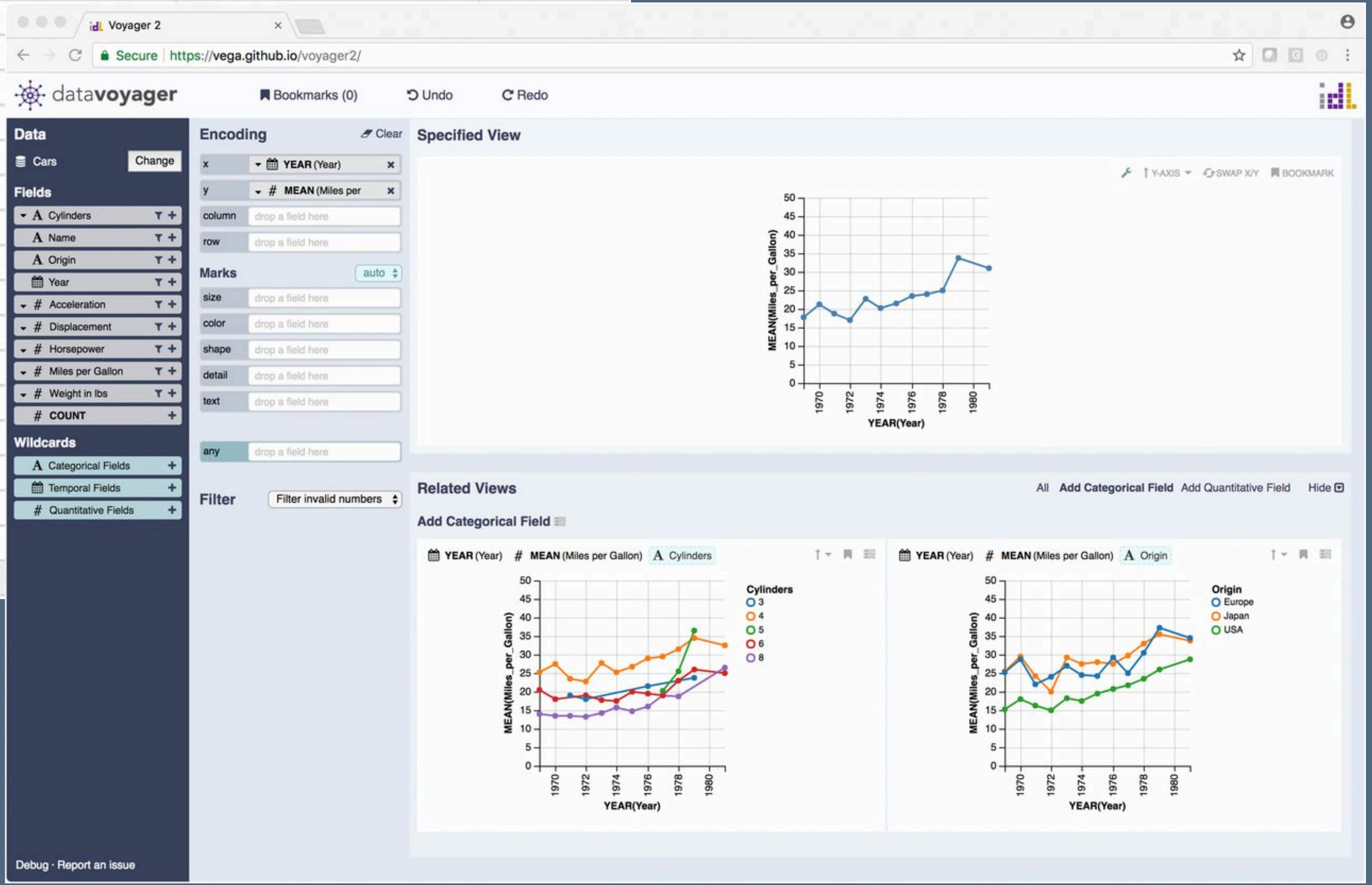
Transform Script (Export)

- Split data repeatedly on newline into rows
- Split data repeatedly on 'tab'
- Delete empty rows
- Promote row 2 to header
- Delete rows where Year contains 'Year'

	# Year	Abc extract	# Population	# Property_crime_rate	# Burgla
1	Reported crime in Alabama	Alabama			
2	2004		4525375	4029.3	987
3	2005		4548327	3900	955.8
4	2006		4599030	3937	968.9
5	2007		4627851		
6	2008		4661900		
7	Reported crime in Alaska	Alaska			
8	2004		657755		
9	2005		663253		
10	2006		670053		
11	2007		683478		
12	2008		686293		
13	Reported crime in Arizona	Arizona			
14	2004		5739879		
15	2005		5953007		
16	2006		6166318		
17	2007		6338755		
18	2008		6500180		
19	Reported crime in Arkansas	Arkansas			
20	2004		2750000		

ROWS: 306

Suggest alternative visualizations



Generalize the user's inputs (selecting text "Alabama") into scripts

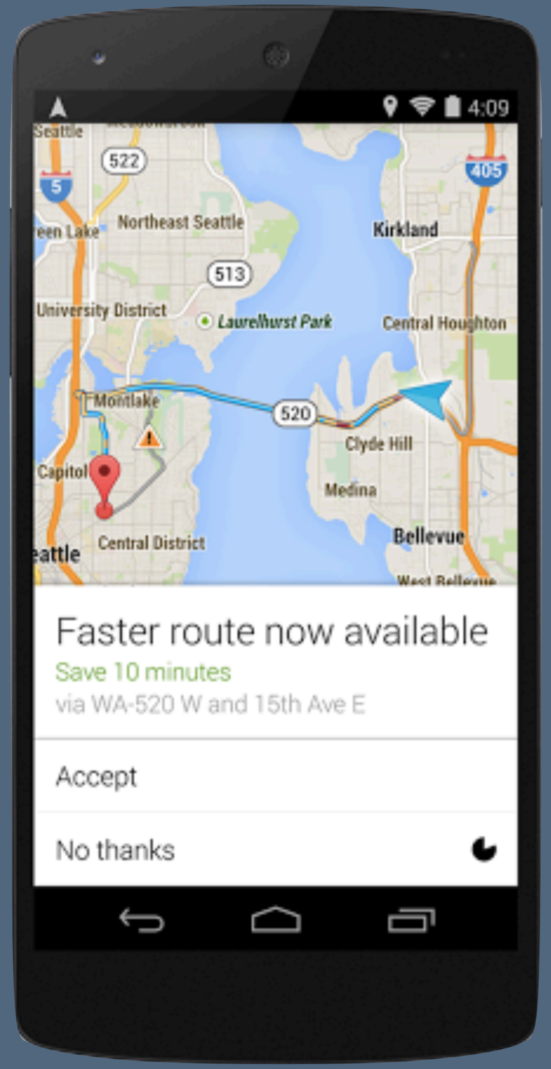
Mixed initiative interaction

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.

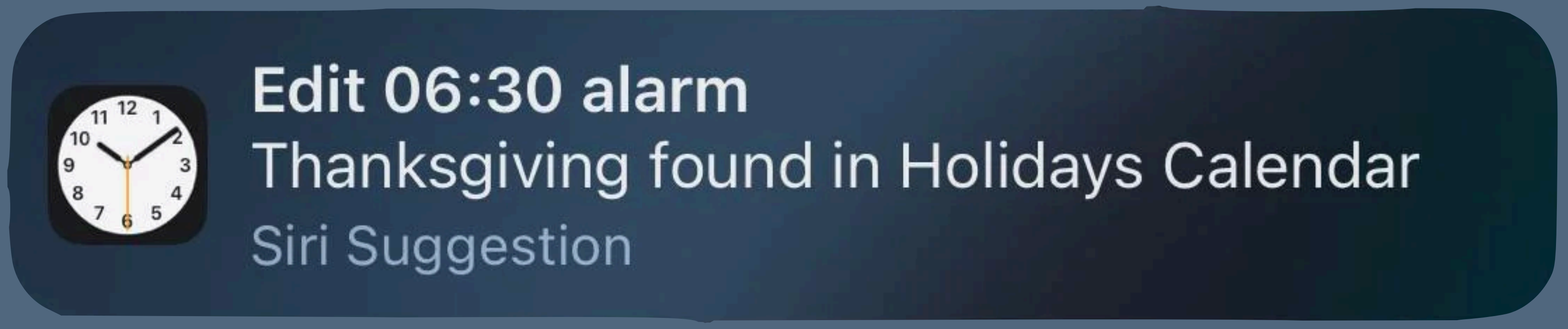
Today, mixed-initiative interaction typically refers to the mode of suggesting an action and letting the user confirm it



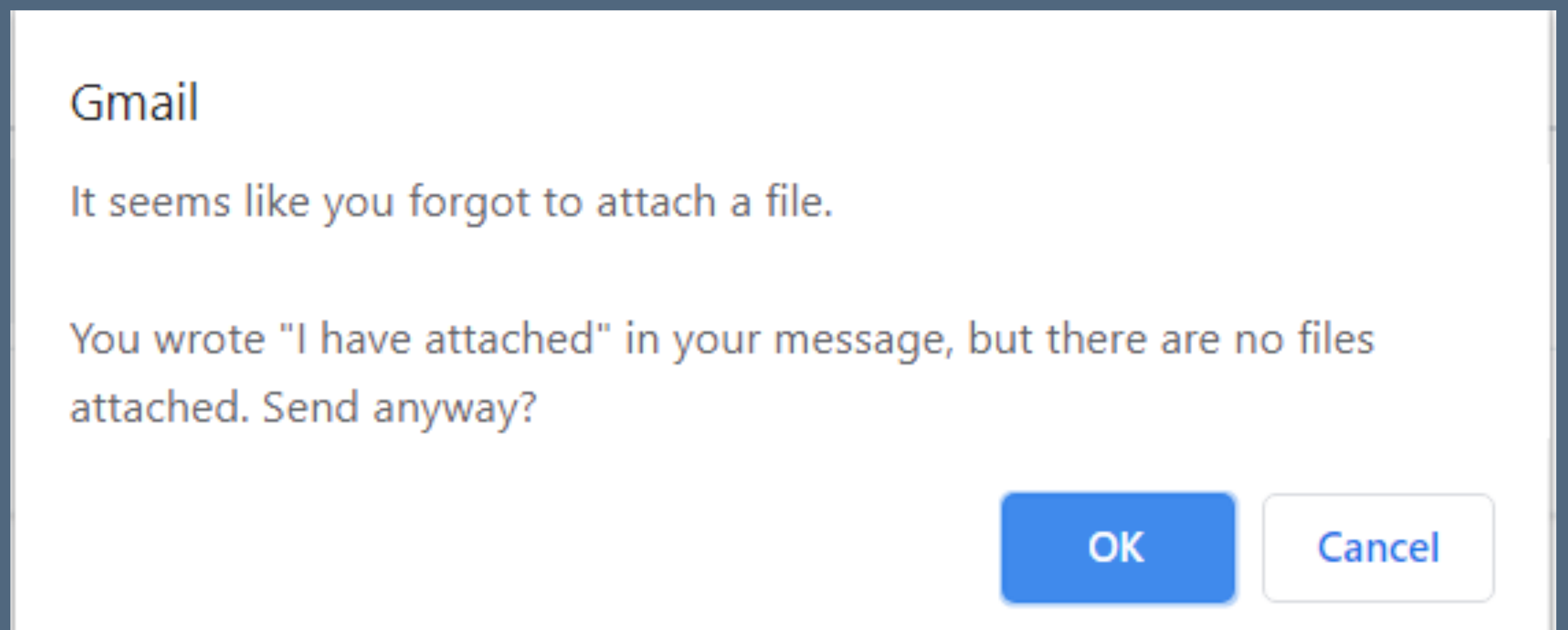
Faster route now available
Save 10 minutes
via WA-520 W and 15th Ave E

Accept

No thanks



Edit 06:30 alarm
Thanksgiving found in Holidays Calendar
Siri Suggestion

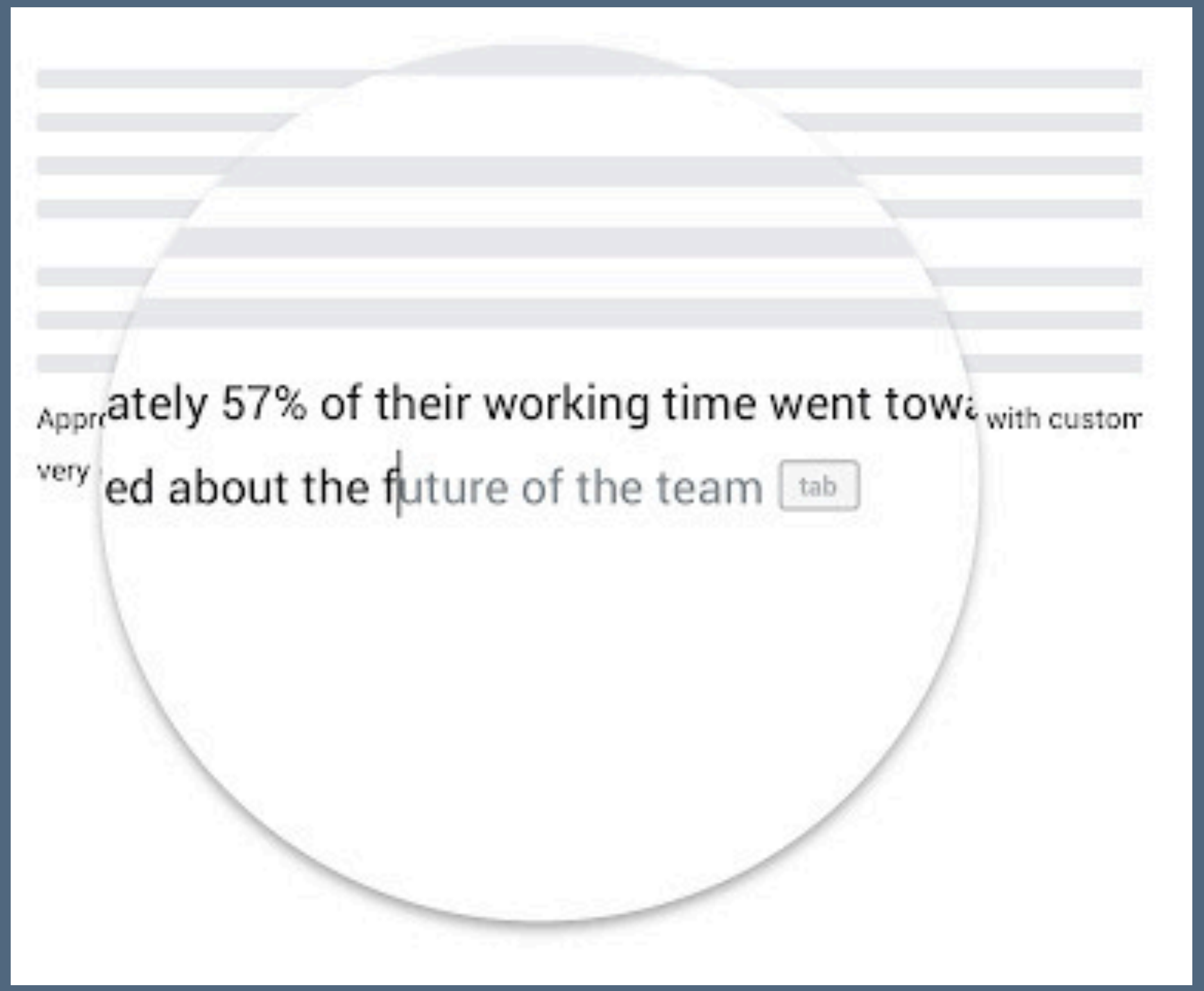


Gmail

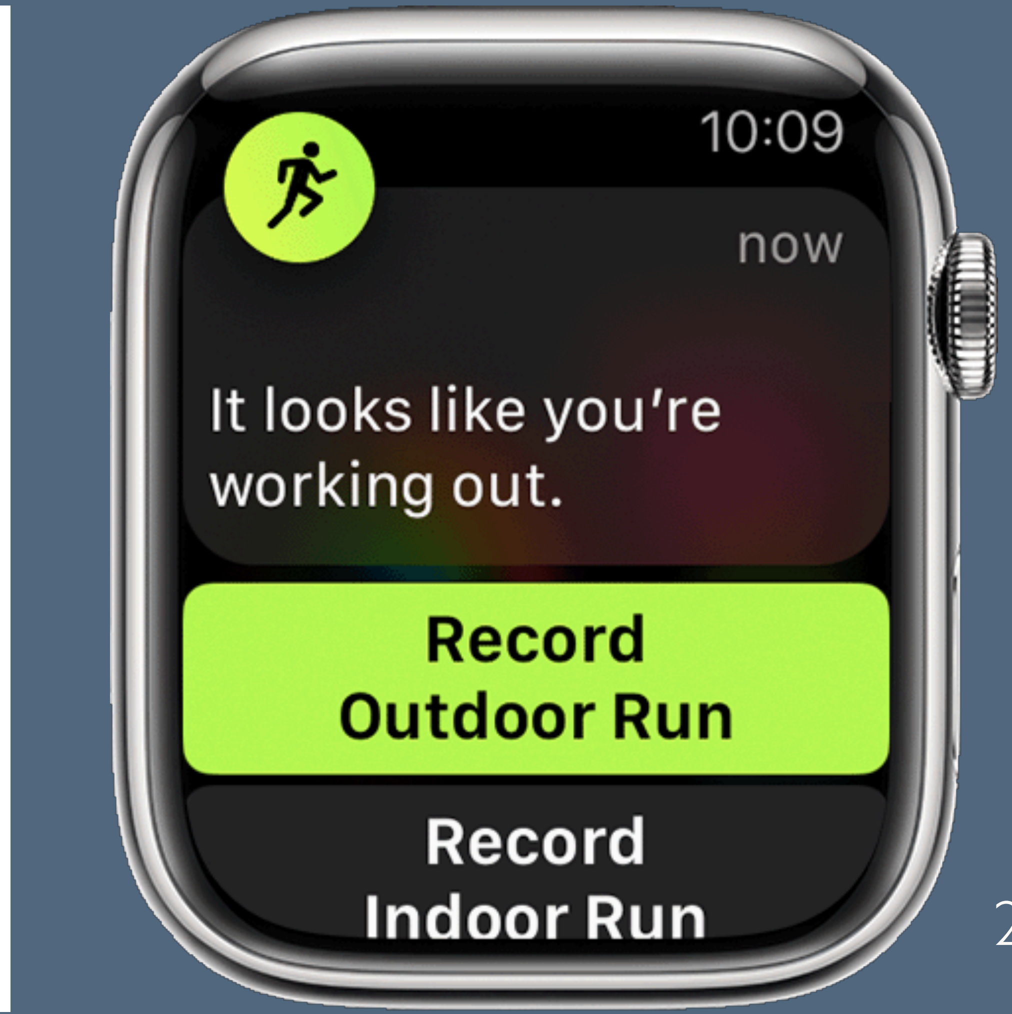
It seems like you forgot to attach a file.

You wrote "I have attached" in your message, but there are no files attached. Send anyway?

OK | Cancel



Appr ately 57% of their working time went towa with custom
very ed about the future of the team tab



10:09
now

It looks like you're working out.

Record Outdoor Run

Record Indoor Run

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?

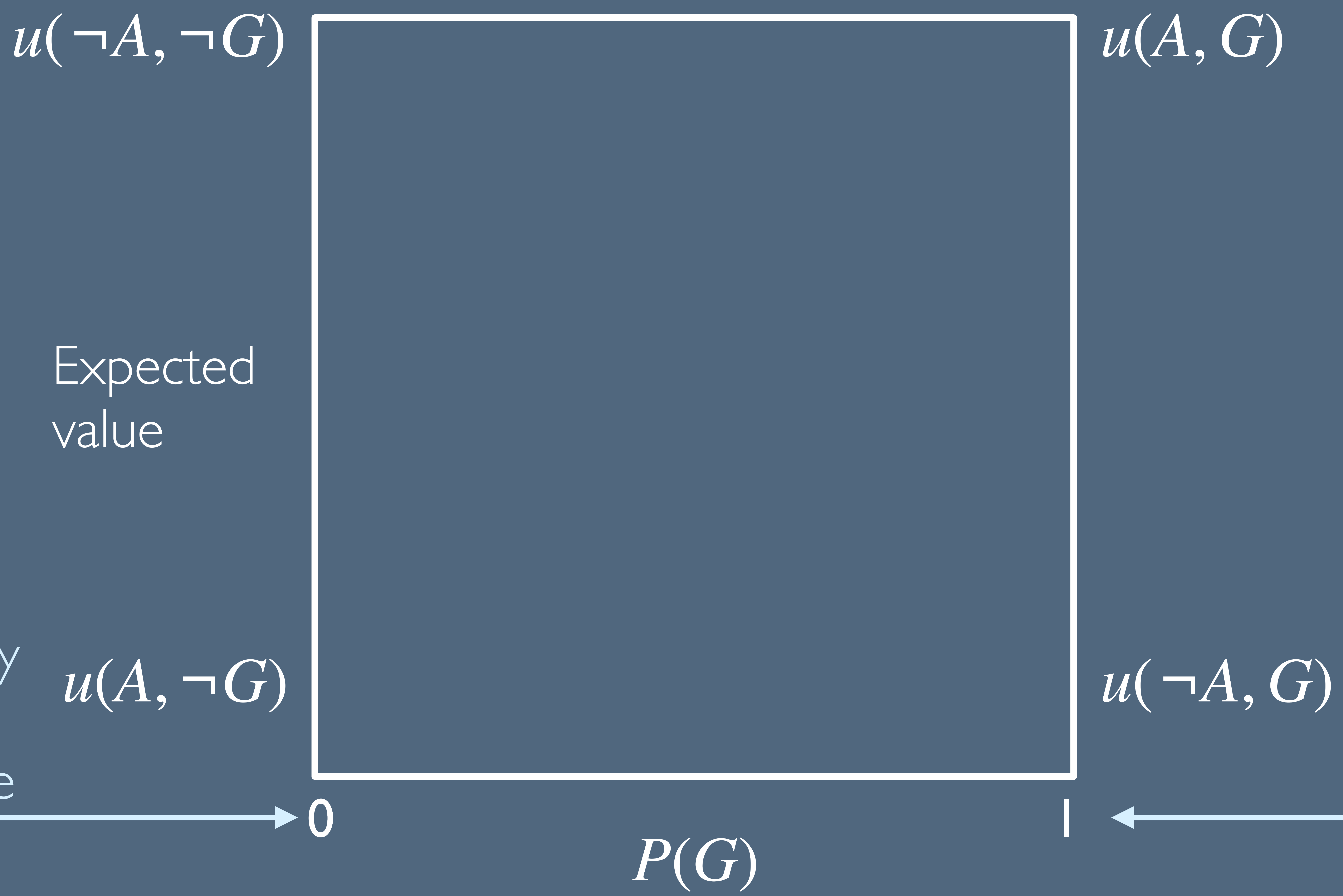
$$P(G) \cdot u(A, G) + P(\neg G) \cdot u(A, \neg G)$$

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)

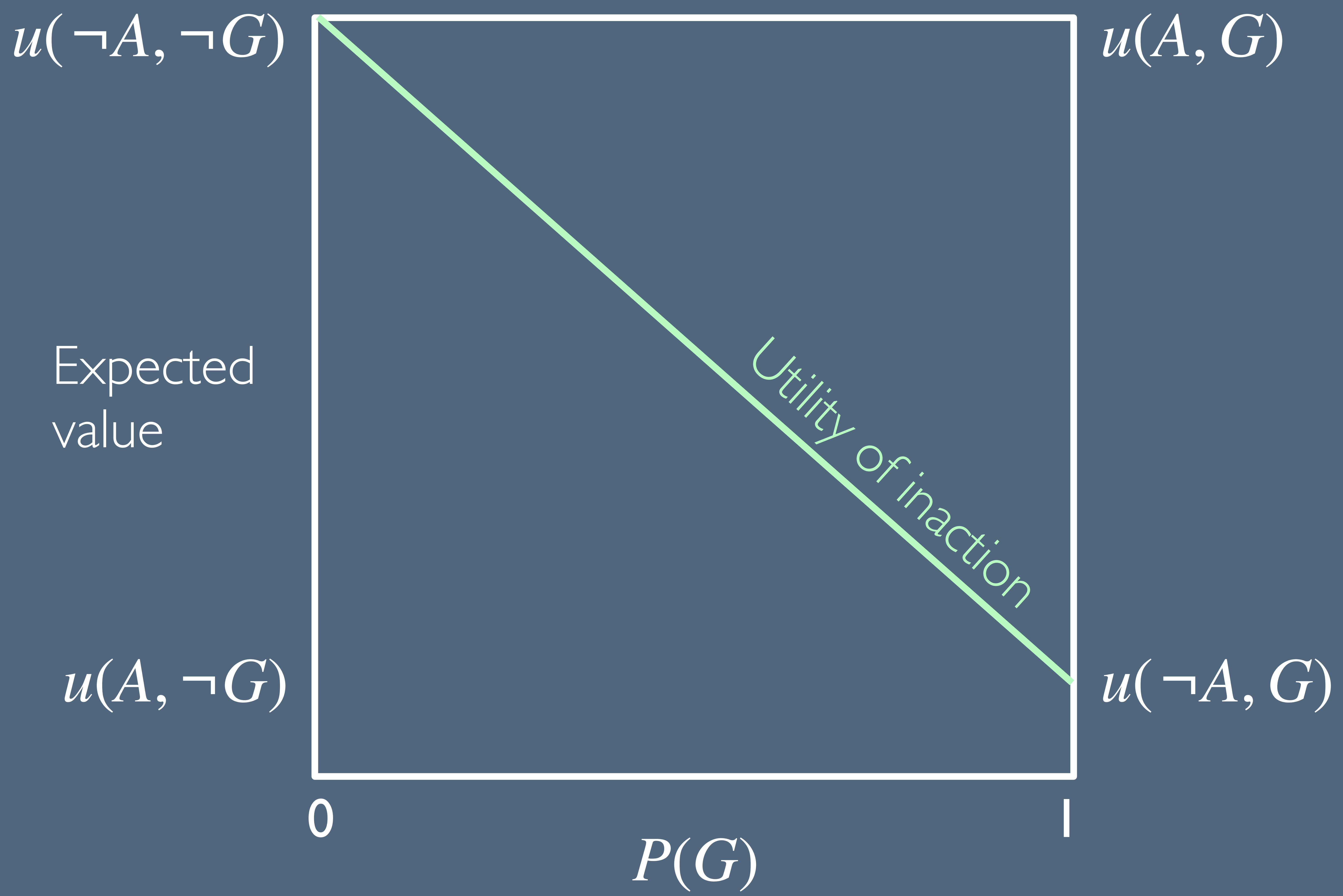
Mixed initiative: visually



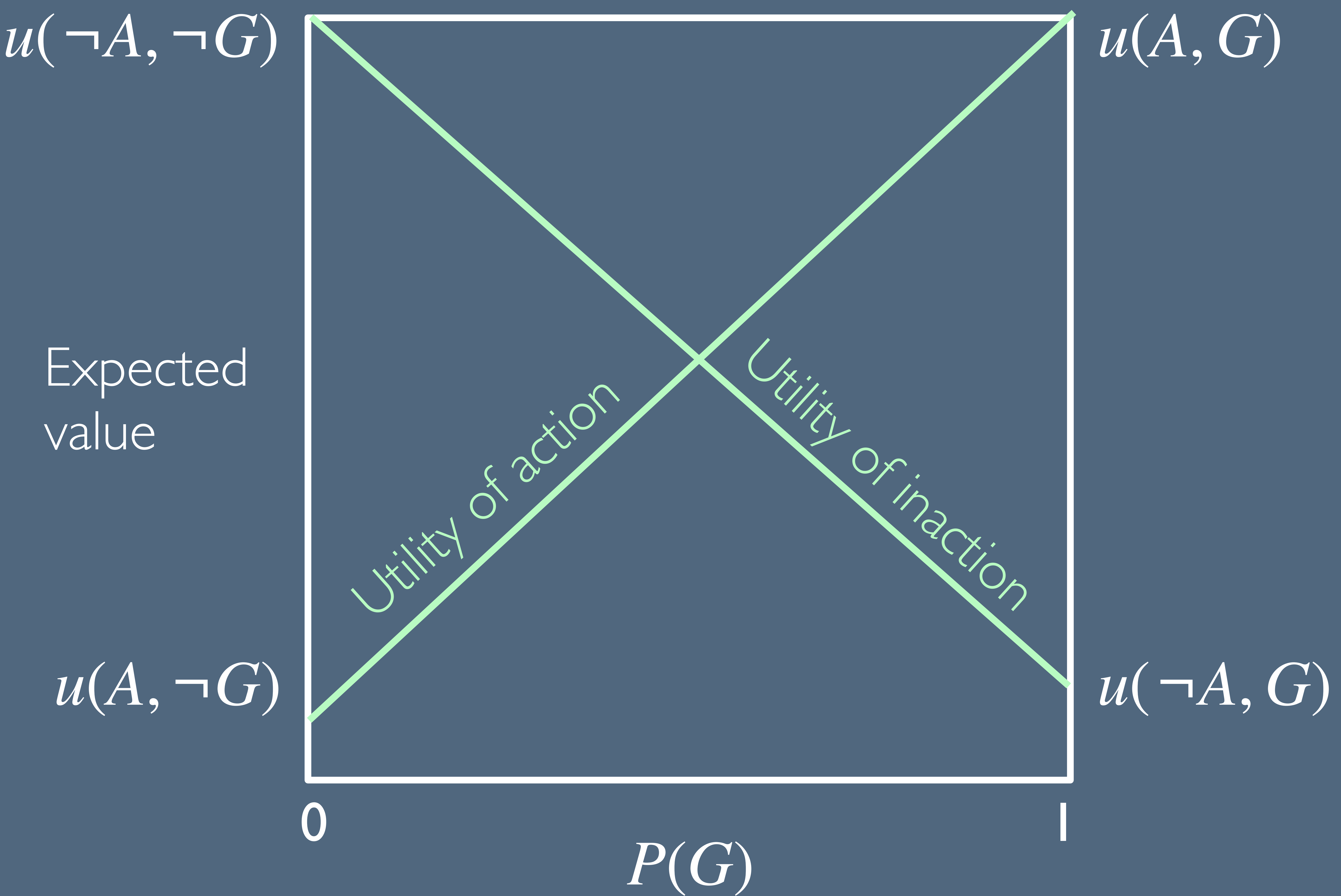
If it's unlikely that the user has the given goal

If it's likely that the user has the given goal

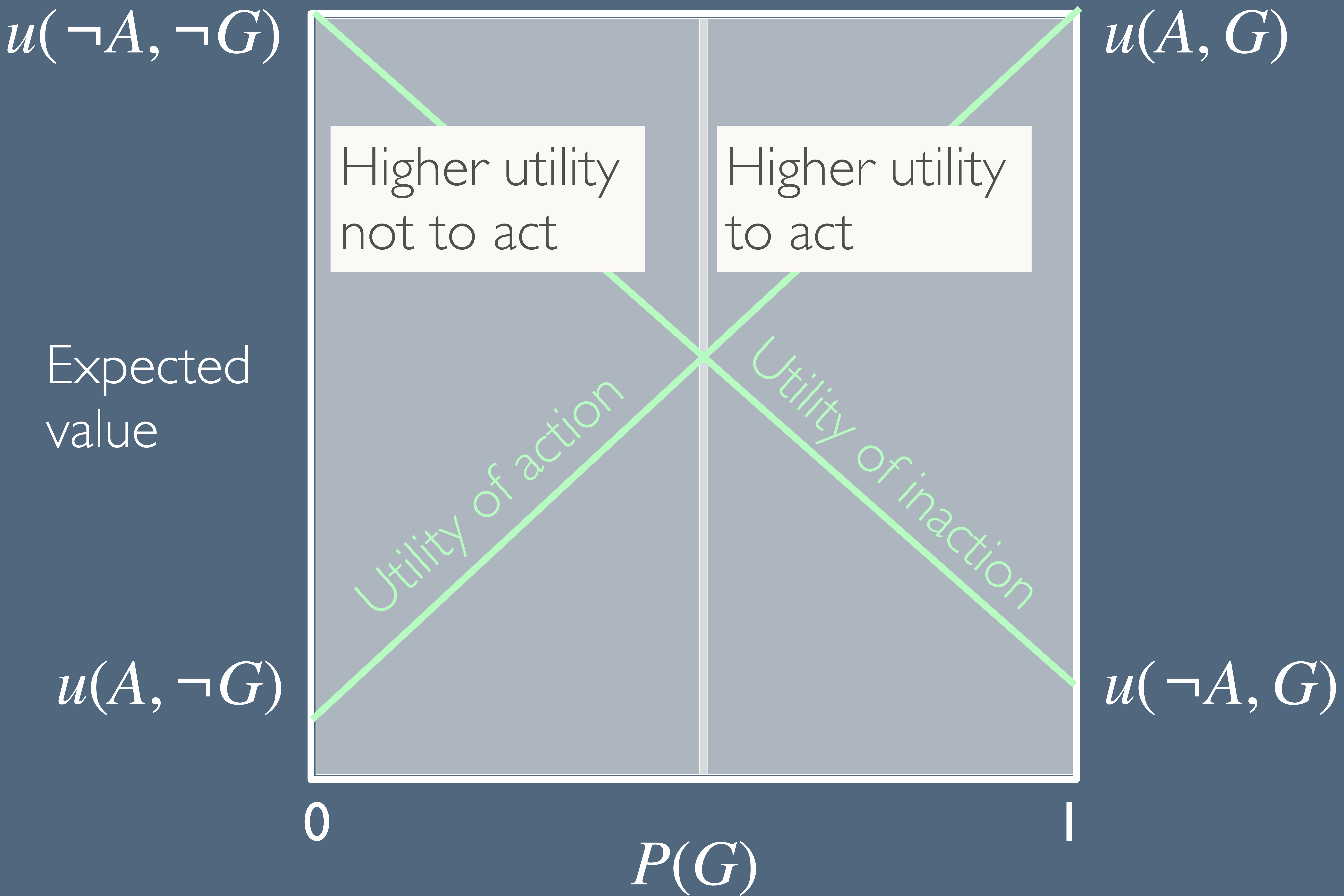
Mixed initiative: visually



Mixed initiative: visually

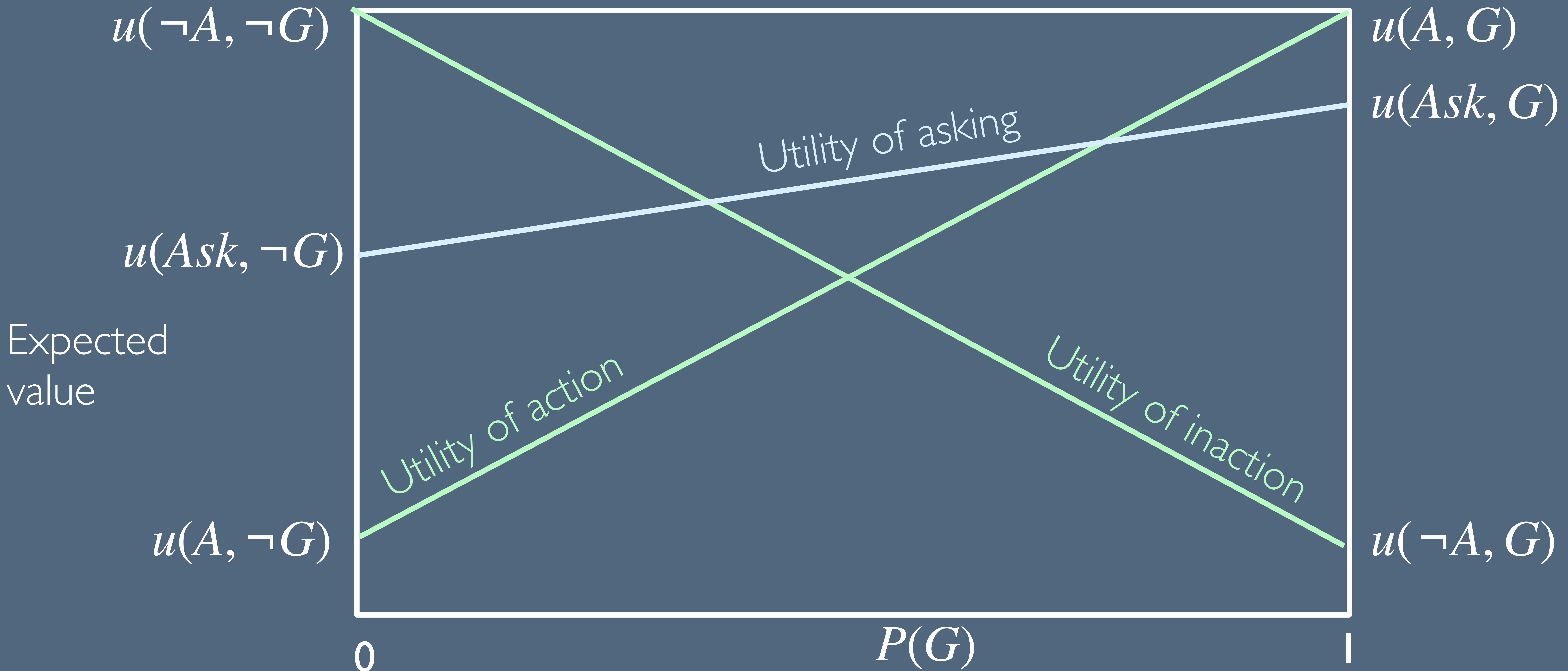


Mixed initiative: visually



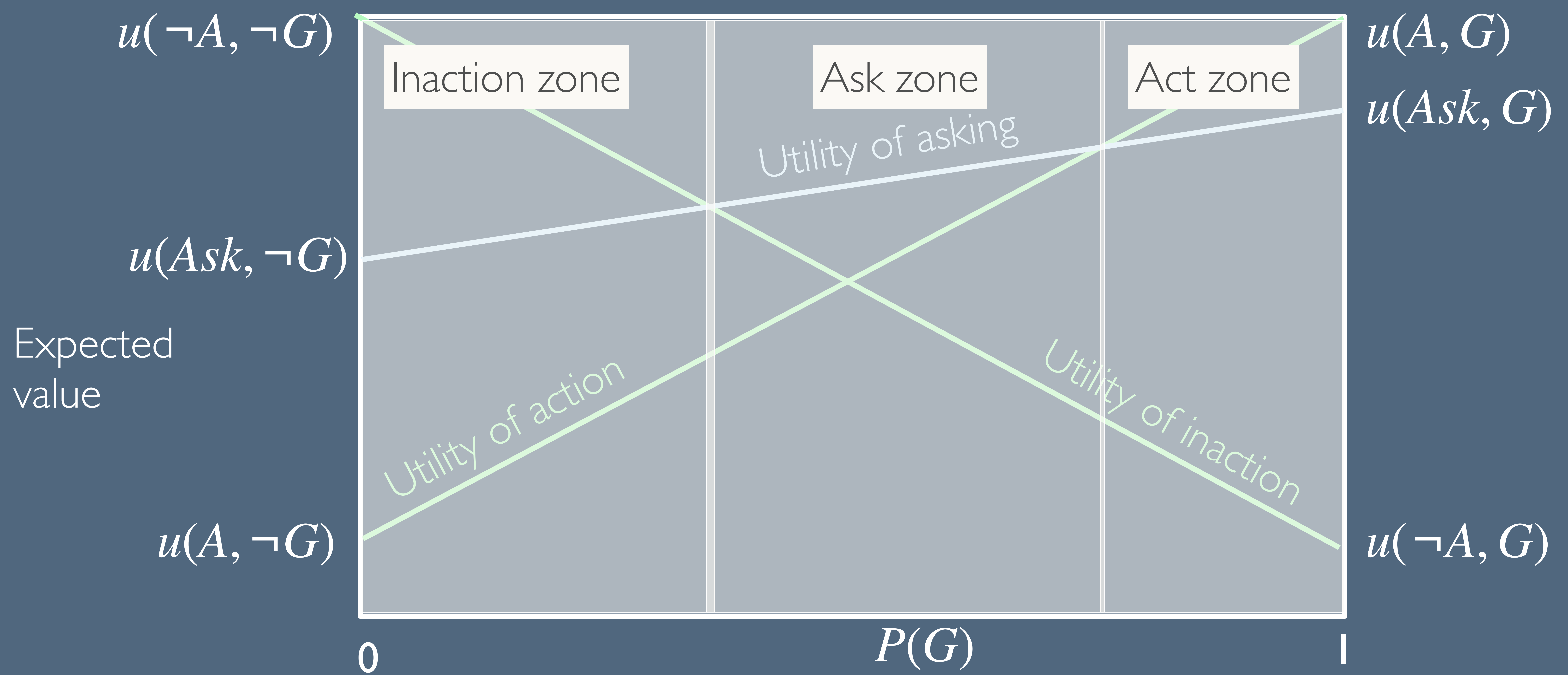
What if we ask the user?

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...

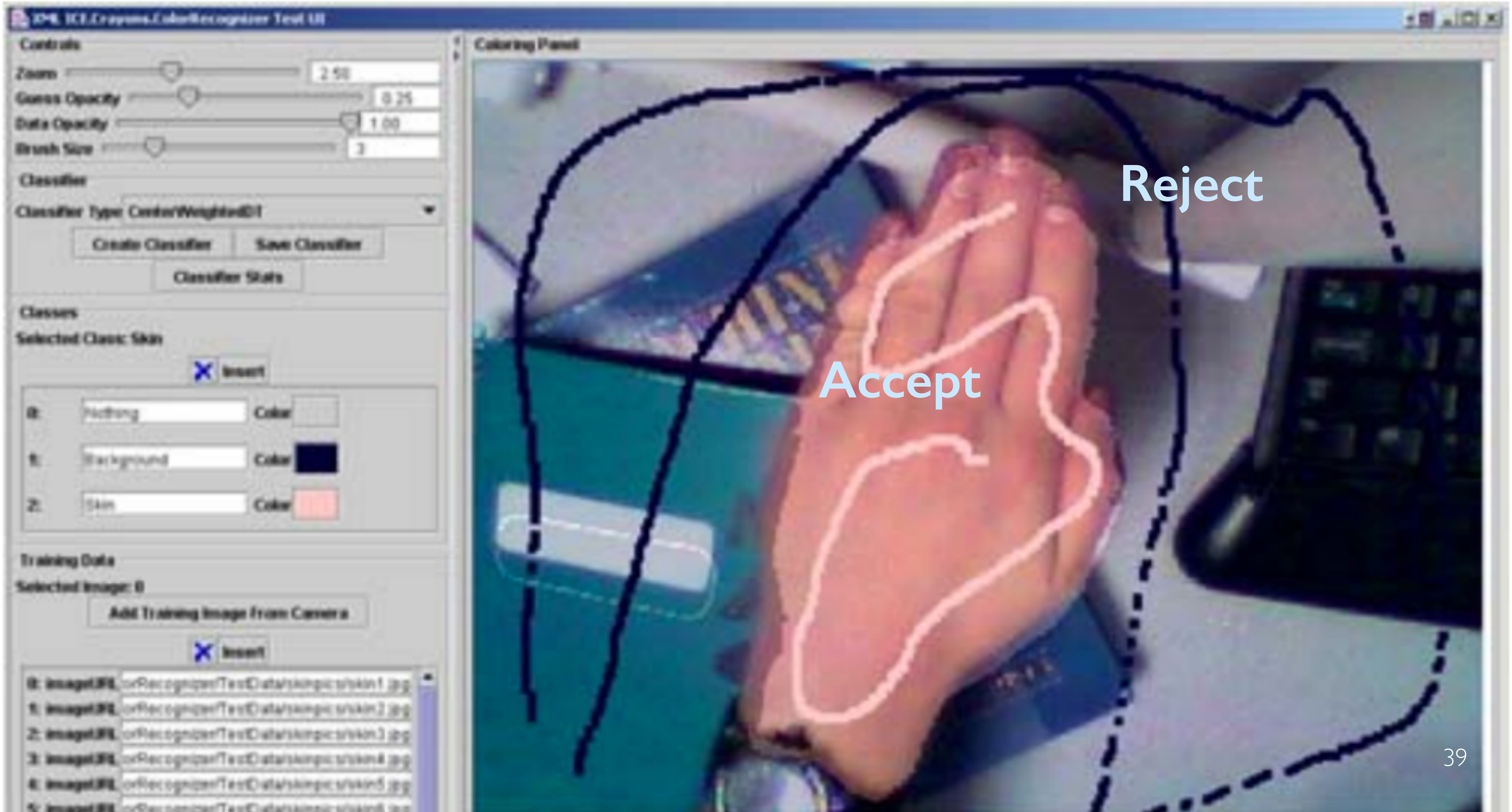
To automatically control entrance to your room to let in possible donors for your Stanford education

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



XML ICE.Crayons.ColorRecognizer Test UI

Controls

Zoom 2.50

Guess Opacity 0.25

Data Opacity 1.00

Brush Size 3

Classifier

Classifier Type: CenterWeightedDT

Classes

Selected Class: Skin

0: Nothing Color

1: Background Color

2: Skin Color

Training Data

Selected Image: 0

0: imageURL orRecognizer/TestData/skinpics/skin1.jpg

1: imageURL orRecognizer/TestData/skinpics/skin2.jpg

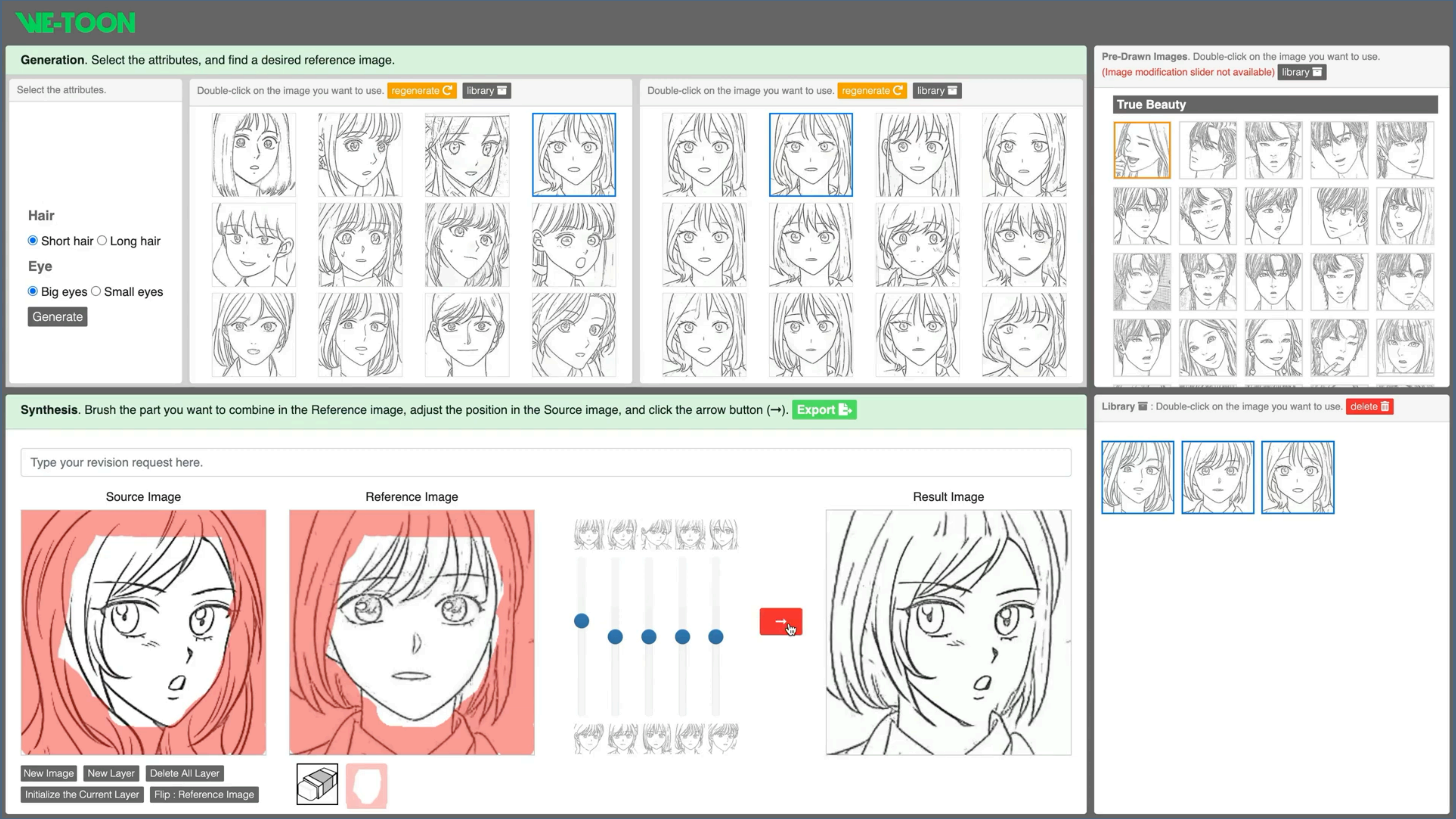
2: imageURL orRecognizer/TestData/skinpics/skin3.jpg

3: imageURL orRecognizer/TestData/skinpics/skin4.jpg

4: imageURL orRecognizer/TestData/skinpics/skin5.jpg

5: imageURL orRecognizer/TestData/skinpics/skin6.jpg

Frontier: image editing through demonstration



Generation. Select the attributes, and find a desired reference image.

Select the attributes.

Hair

● Short hair ○ Long hair

Eye

● Big eyes ○ Small eyes

Double-click on the image you want to use.

Pre-Drawn Images. Double-click on the image you want to use. (Image modification slider not available)

Synthesis. Brush the part you want to combine in the Reference image, adjust the position in the Source image, and click the arrow button (→).

Type your revision request here.

Library : Double-click on the image you want to use.

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

Interactive training

[Fogarty et al. 2008]

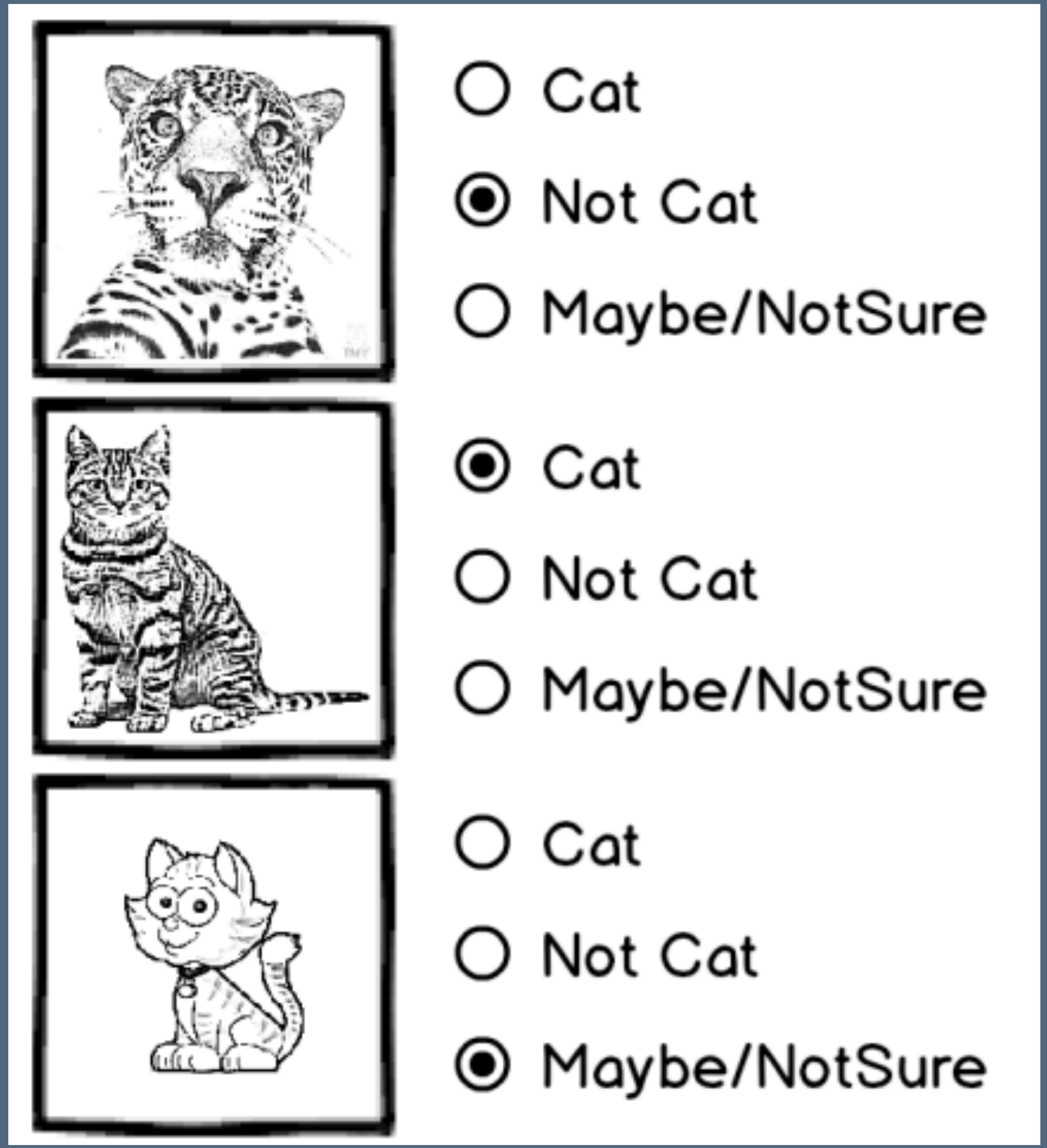
Allow users to keep training and re-training by drag-dropping instances into positive and negative classes as they go

Image Search for "Mountain"
Creating a "Scenic" Rule
Apply the "Scenic" Rule
CueFlik
mountain
Active Rules
Available Rules
Scenic
Editing Rule:
Images You Said This Rule Should Match
Images You Said This Rule Should Reject

Revising your training as you go

[Chang, Amershi and Kamar 2017]

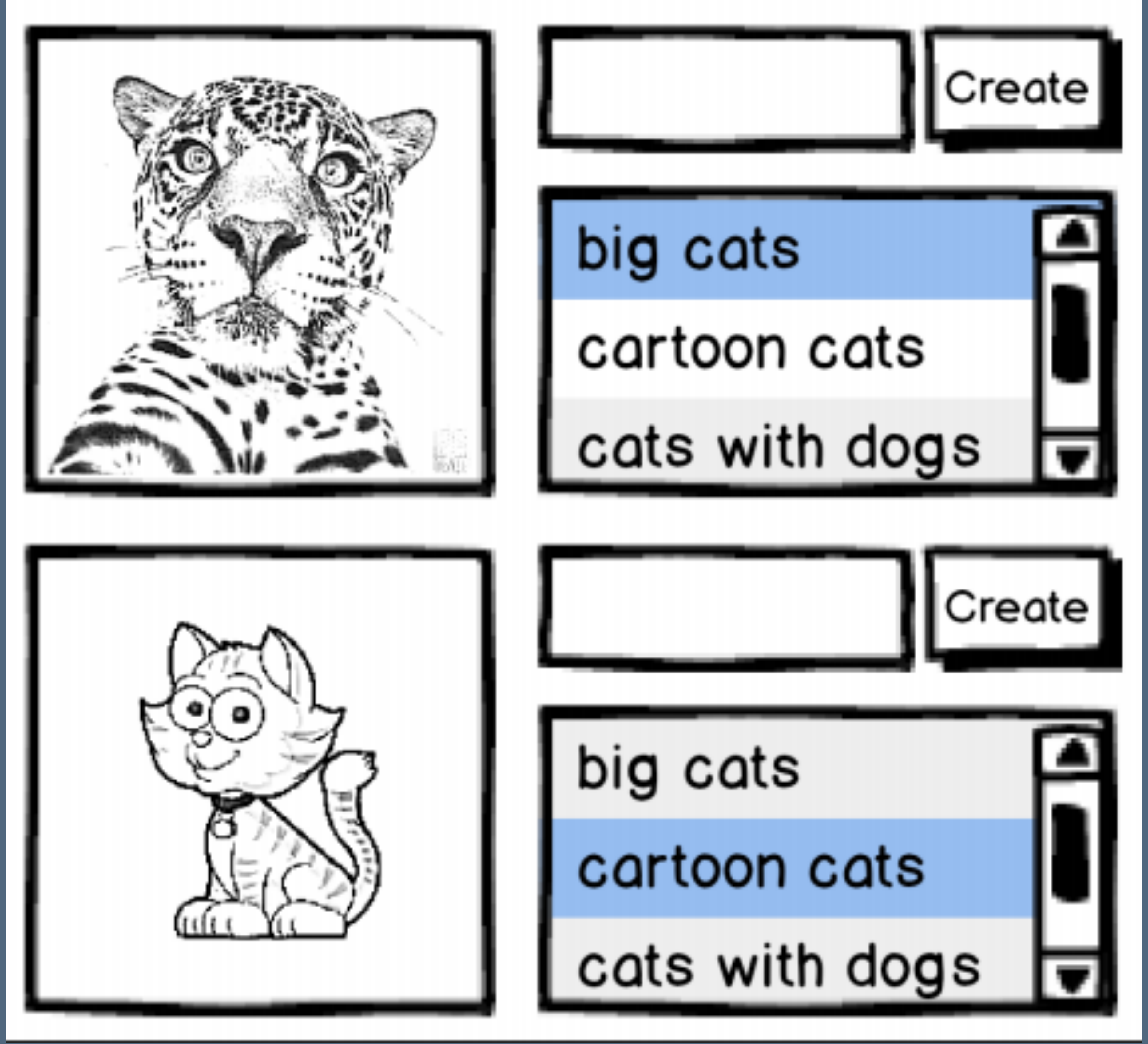
Facilitate concept evolution through a “could be” category that allows clustering into subcategories you can change labels for



○ Cat
● Not Cat
○ Maybe/NotSure

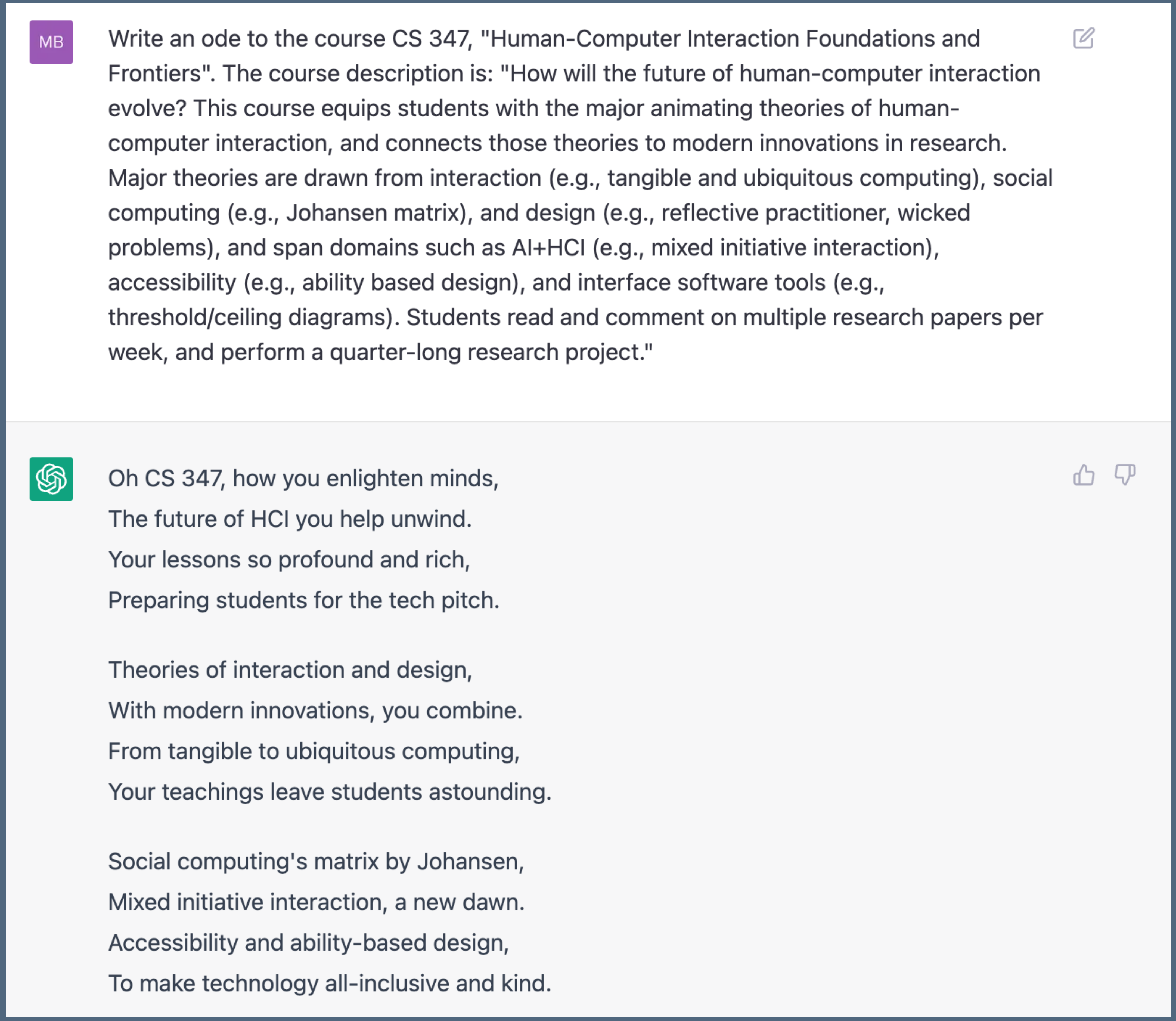
● Cat
○ Not Cat
○ Maybe/NotSure

○ Cat
○ Not Cat
● Maybe/NotSure



More recently: prompting

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



MB

Write an ode to the course CS 347, "Human-Computer Interaction Foundations and Frontiers". The course description is: "How will the future of human-computer interaction evolve? This course equips students with the major animating theories of human-computer interaction, and connects those theories to modern innovations in research. Major theories are drawn from interaction (e.g., tangible and ubiquitous computing), social computing (e.g., Johansen matrix), and design (e.g., reflective practitioner, wicked problems), and span domains such as AI+HCI (e.g., mixed initiative interaction), accessibility (e.g., ability based design), and interface software tools (e.g., threshold/ceiling diagrams). Students read and comment on multiple research papers per week, and perform a quarter-long research project."

Oh CS 347, how you enlighten minds,
The future of HCI you help unwind.
Your lessons so profound and rich,
Preparing students for the tech pitch.

Theories of interaction and design,
With modern innovations, you combine.
From tangible to ubiquitous computing,
Your teachings leave students astounding.

Social computing's matrix by Johansen,
Mixed initiative interaction, a new dawn.
Accessibility and ability-based design,
To make technology all-inclusive and kind.

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?

The challenge of designing with AI

YOU READ THIS

Why AI is difficult to design

[Yang et al. 2020]

How do we know what AI can and cannot do, and how it will err?

How do we engage in rapid prototyping of AI-powered systems?

How do we control the unpredictable output of the AI?

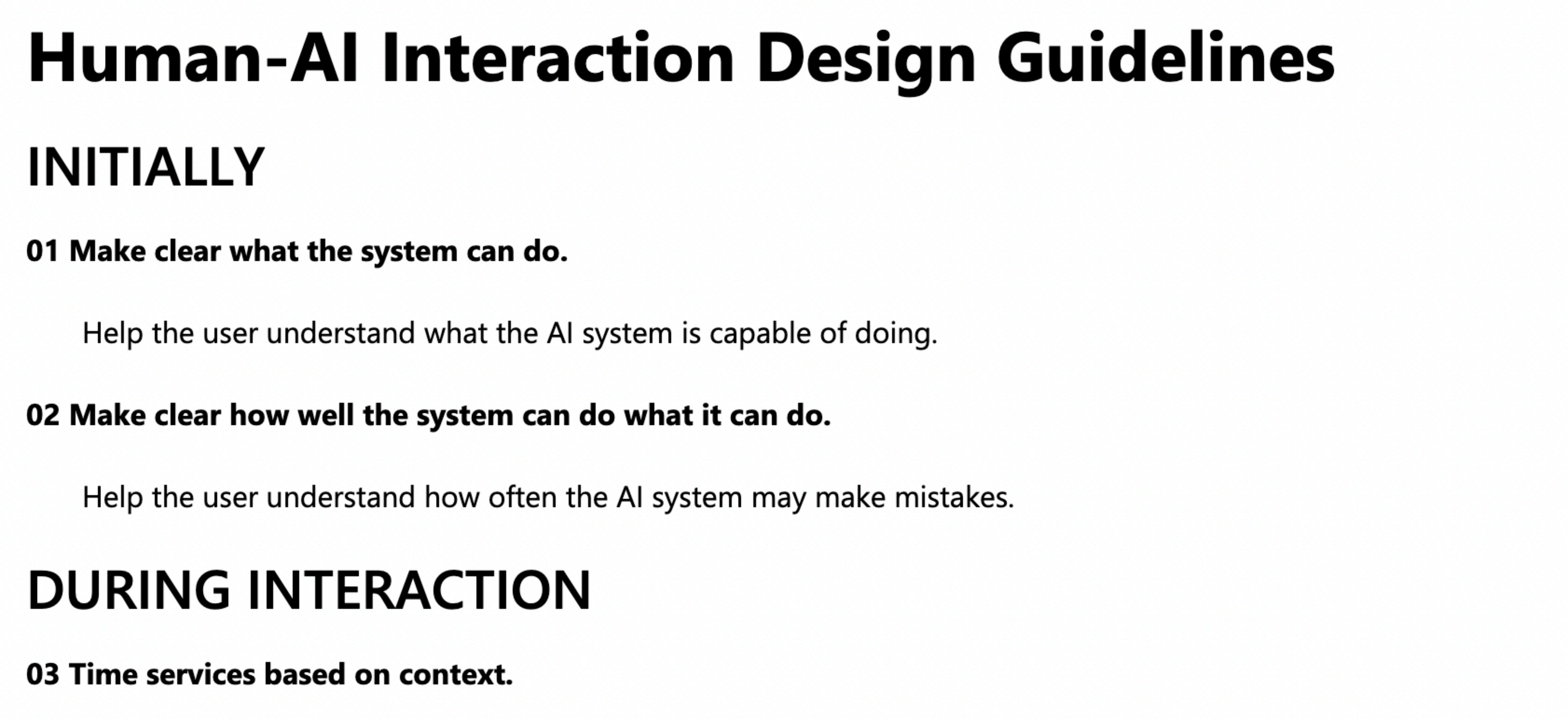
I would add:

> We are risk averse and will avoid AI-powered interactions once we stumble into one of their limits: **algorithm aversion**.
>
> If "Alexa, play a reggae song by Beyoncé" returns the wrong thing, or your text message dictation errs, you back off to simpler interactions

Human-AI design guidelines

[Amershi et al. 2019]

What guidelines, similar to Nielsen's heuristic evaluation principles, ought to apply for human-AI interaction design?



Human-AI Interaction Design Guidelines

INITIALLY

01 Make clear what the system can do.

Help the user understand what the AI system is capable of doing.

02 Make clear how well the system can do what it can do.

Help the user understand how often the AI system may make mistakes.

DURING INTERACTION

03 Time services based on context.

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 assessment of the benefit

End users and designers seek to work with these tools

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.

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.