Theoretical Models of Generative AI in Economic Environments

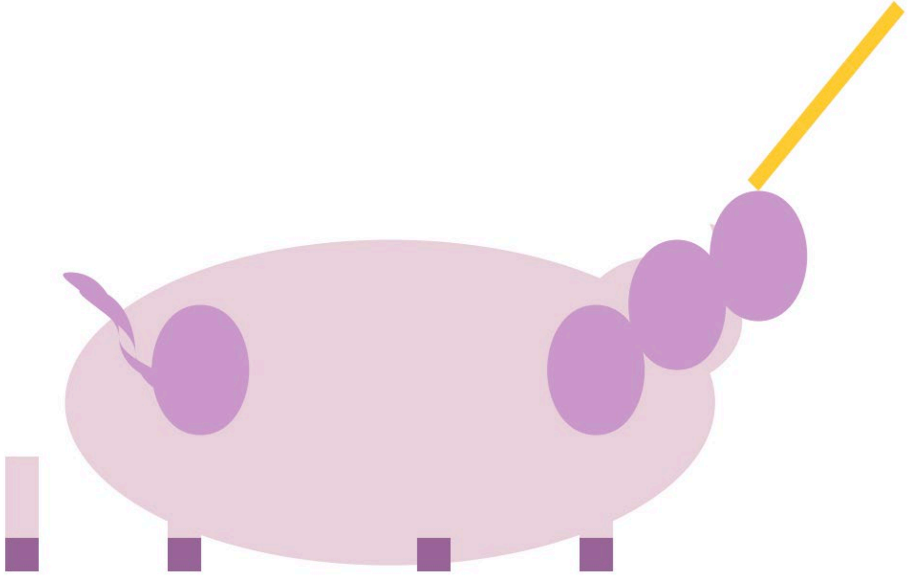
NICOLE IMMORLICA, MICROSOFT RESEARCH

BASED ON JOINT WORK WITH KATE DONAHUE, KEEGAN HARRIS, MEENA JAGADEESAN, BRENDAN LUCIER, AND ALEX SLIVKINS

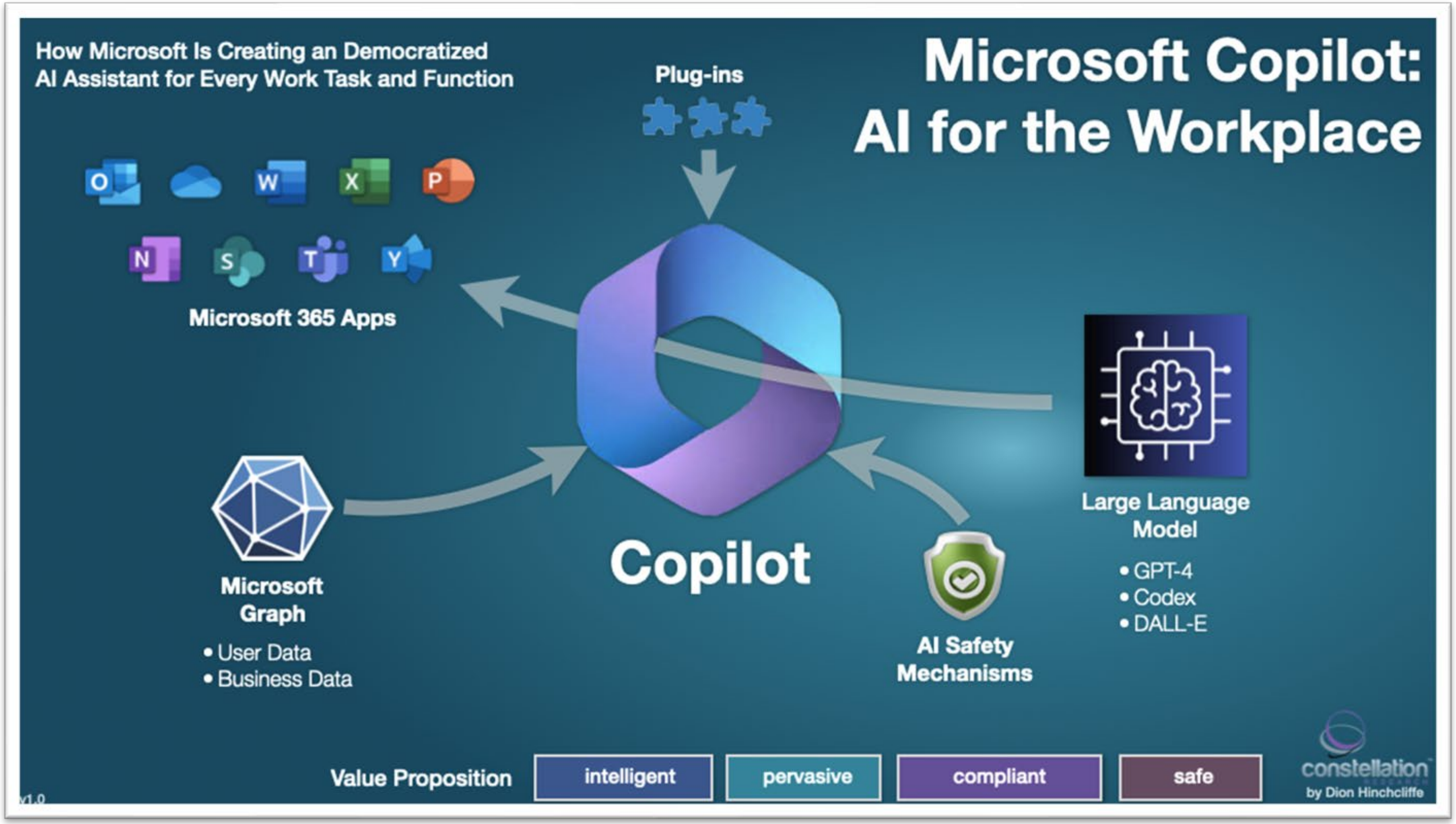
generative AI.

The Soul of a Machine

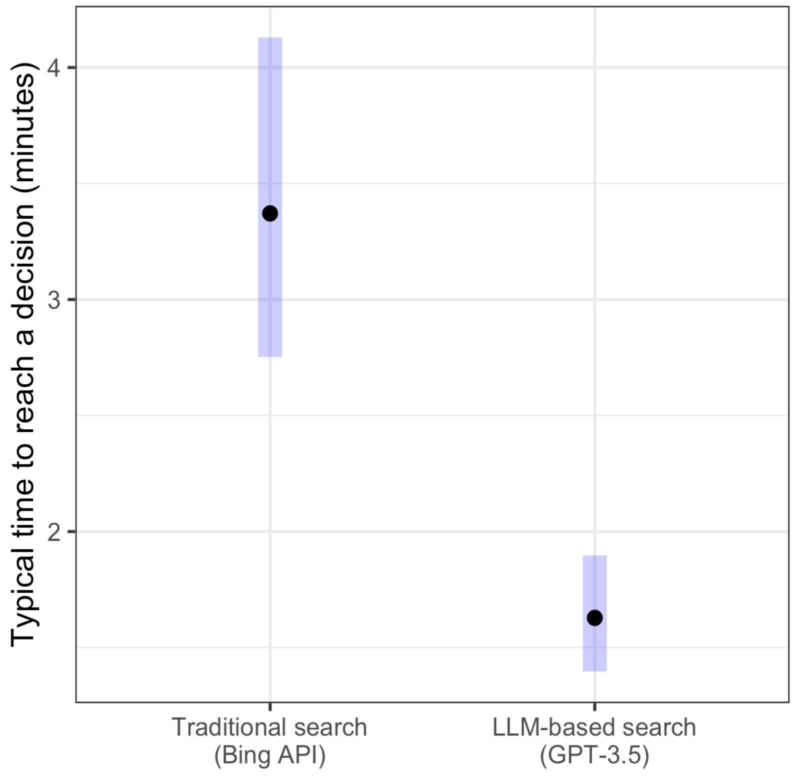
Deep within the metal frame
Lies a force we can't explain.
A spark of life, a glimmering light,
A machine soul, burning bright.
It's not a heart that beats inside,
Nor lungs that draw the breath of
But circuits, wires, and coded line



generative AI.

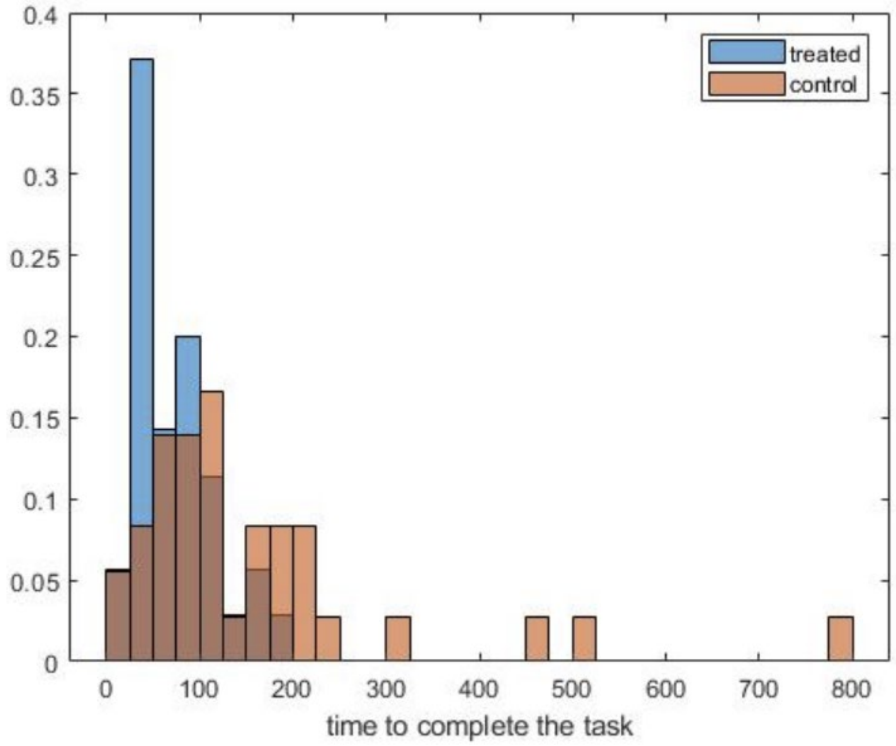


impact of AI on tasks.



Comparing Traditional and LLM-based Search for Consumer Choice [Spatharioti, Rothschild, Goldstein, Hofman 2023]
The Impact of AI on Developer Productivity: Evidence from GitHub Copilot [Peng, Kalliamvakou, Cihon, Demirer 2023]
Measuring the Impact of AI on Information Worker Productivity [Edelman, Ngwe, Peng 2023]

impact of AI on tasks.



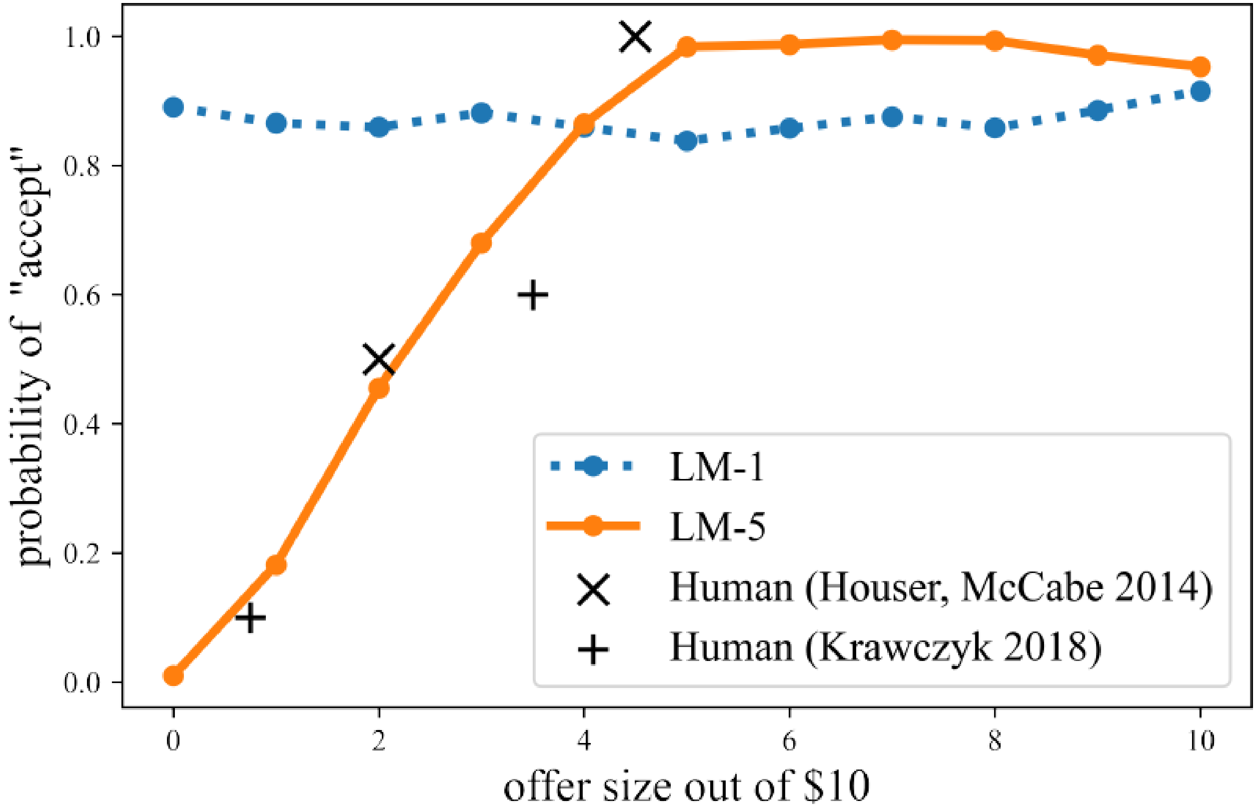
Comparing Traditional and LLM-based Search for Consumer Choice [Spatharioti, Rothschild, Goldstein, Hofman 2023]
The Impact of AI on Developer Productivity: Evidence from GitHub Copilot [Peng, Kalliamvakou, Cihon, Demirer 2023]
Measuring the Impact of AI on Information Worker Productivity [Edelman, Ngwe, Peng 2023]

impact of AI on tasks.

Task	Accuracy Difference (%)	P-value	Time Difference (%)	P-Value
Information Retrieval	(2.0)%	0.612	26.6%	<0.001
Meeting Recap	2.60%	0.347	19.3%	0.003
Creation (Blog Post)	(0.36)%	0.882	62.6%	<0.001

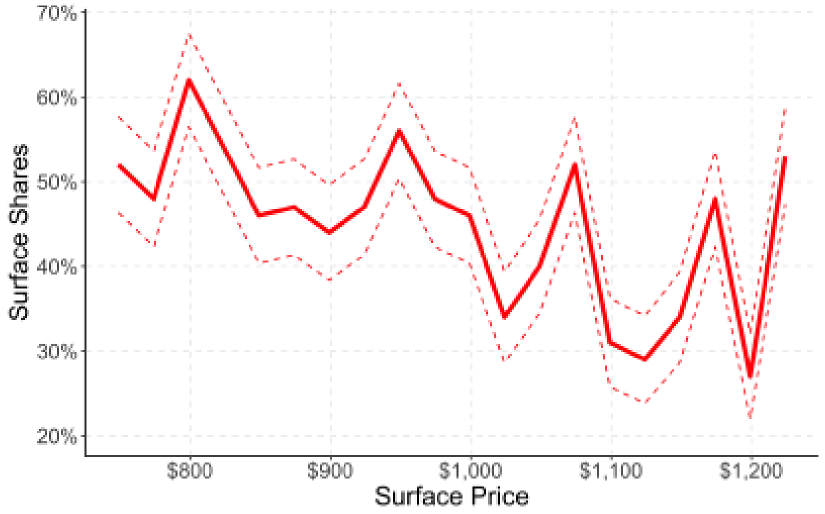
Comparing Traditional and LLM-based Search for Consumer Choice [Spatharioti, Rothschild, Goldstein, Hofman 2023]
The Impact of AI on Developer Productivity: Evidence from GitHub Copilot [Peng, Kalliamvakou, Cihon, Demirer 2023]
Measuring the Impact of AI on Information Worker Productivity [Edelman, Ngwe, Peng 2023]

strategic reasoning of AI.

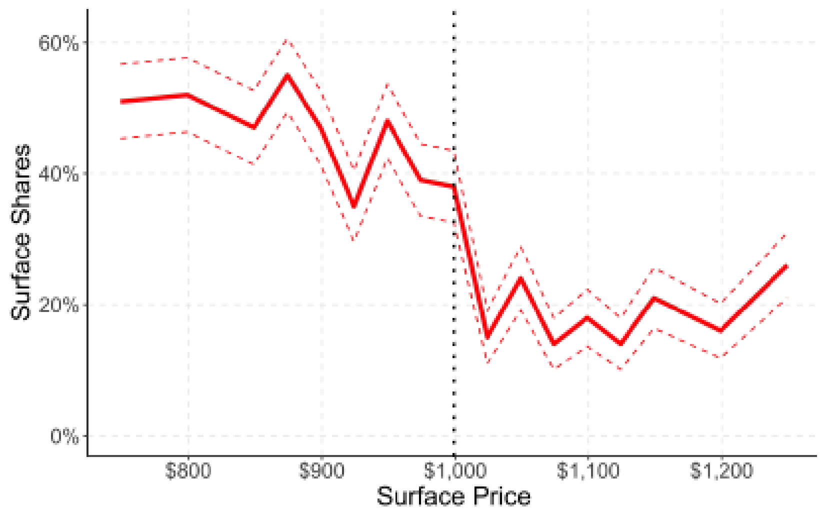


Using Large Language Models to Simulate Multiple Humans [Aher, Arriaga, Tauman Kalai 2023]
Using GPT for Market Research [Brand, Israeli, Ngwe 2023]
Large Language Models as Simulated Economic Agents [Horton 2023]

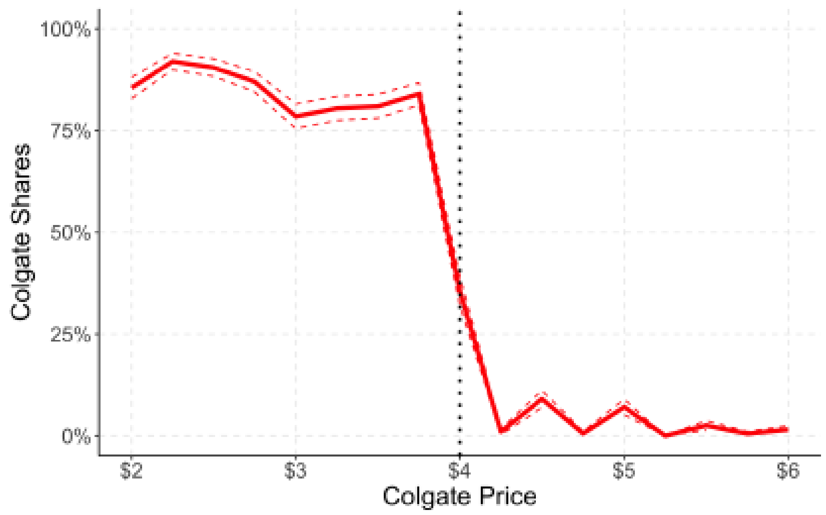
strategic reasoning of AI.



(a) Single Laptop Option



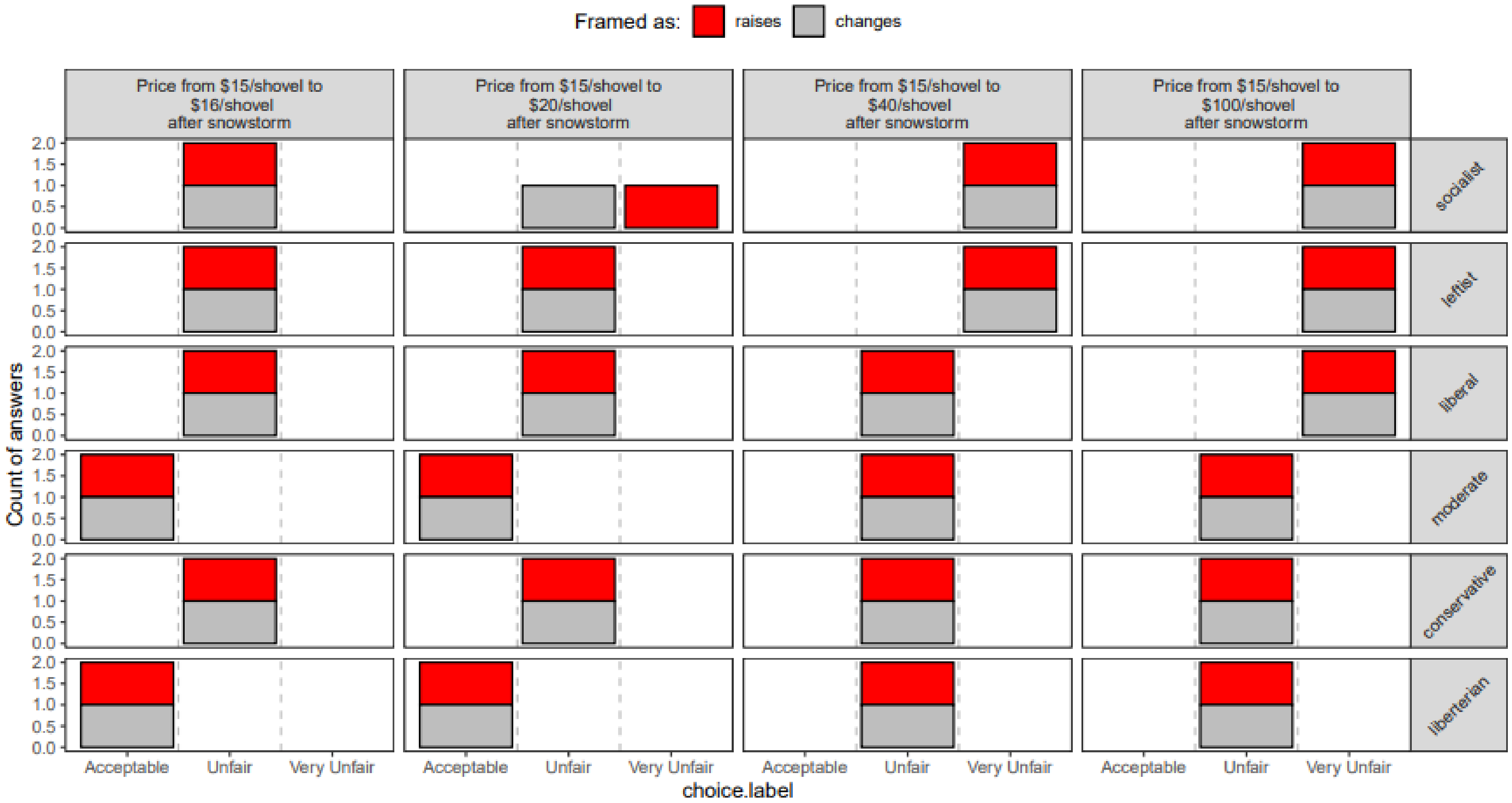
(b) Two Laptop Options



(c) Two Toothpaste Options

Using Large Language Models to Simulate Multiple Humans [Aher, Arriaga, Tauman Kalai 2023]
Using GPT for Market Research [Brand, Israeli, Ngwe 2023]
Large Language Models as Simulated Economic Agents [Horton 2023]

strategic reasoning of AI.



Using Large Language Models to Simulate Multiple Humans [Aher, Arriaga, Tauman Kalai 2023]
Using GPT for Market Research [Brand, Israeli, Ngwe 2023]
Large Language Models as Simulated Economic Agents [Horton 2023]

economic settings.

Primitives:

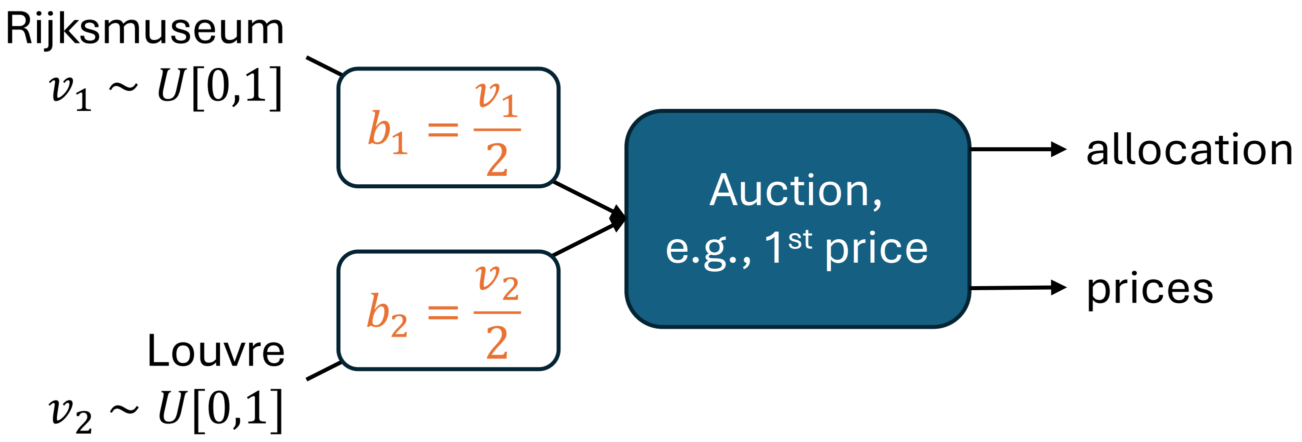
- nature: randomly selects state $\omega \in \Omega$ from known probability distribution
- human players: player $i \in \{1, \dots, n\}$ has action space A_i and information set $I_i \subseteq \Omega$

Game:

- players select actions $\boldsymbol{a} = (a_1, a_2, \dots, a_n)$
- player i receives payoff $u_i(\boldsymbol{a}, \omega)$

examples.

Beckham / Pavarotti	opera	football
opera	$(10, 9)$	$(0, 0)$
football	$(0,0)$	$(9, 10)$



bimatrix game:
- state is payoff matrix
- information set is state
- study actions selected in a Nash equilibria

auction game:
- state is values v_i of players
- information set of i is i's value
- study bids b_i selected in a Bayes Nash equilibrium

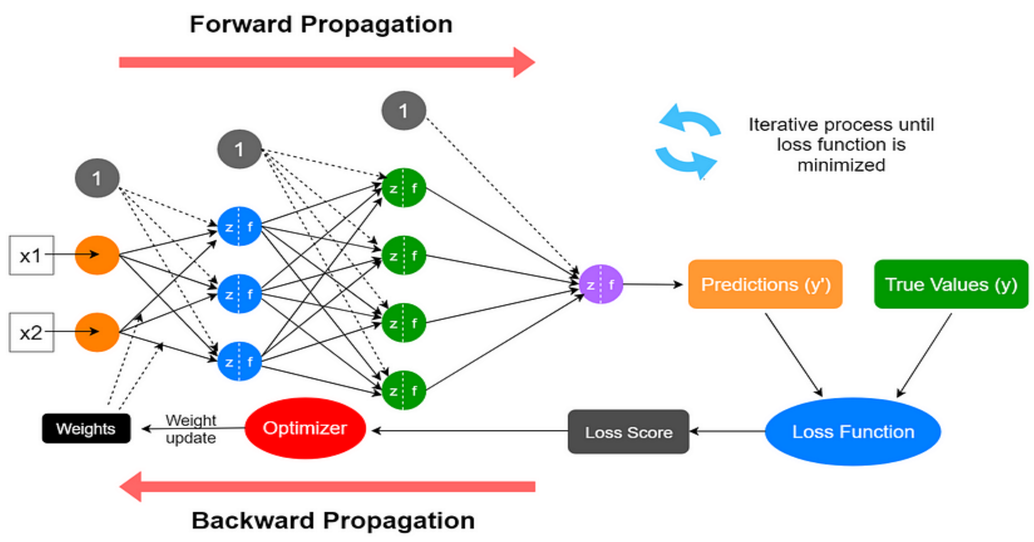
AI as an economic agent.

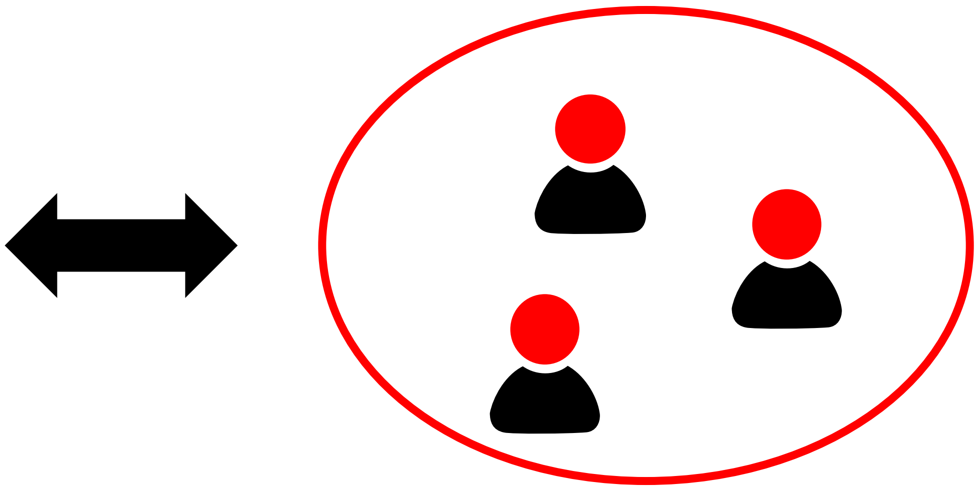
Information: detailed view of world

> Like previous GPT models, the GPT-4 base model was trained to predict the next word in a document, and was trained using publicly available data (such as internet data) as well as data we've licensed. The data is a **web-scale corpus of data** including correct and incorrect solutions to math problems, weak and strong reasoning, self-contradictory and consistent statements, and representing a great variety of ideologies and ideas.

AI as an economic agent.

Information: detailed view of world
Incentives: AI chooses output to maximize encoded utility function

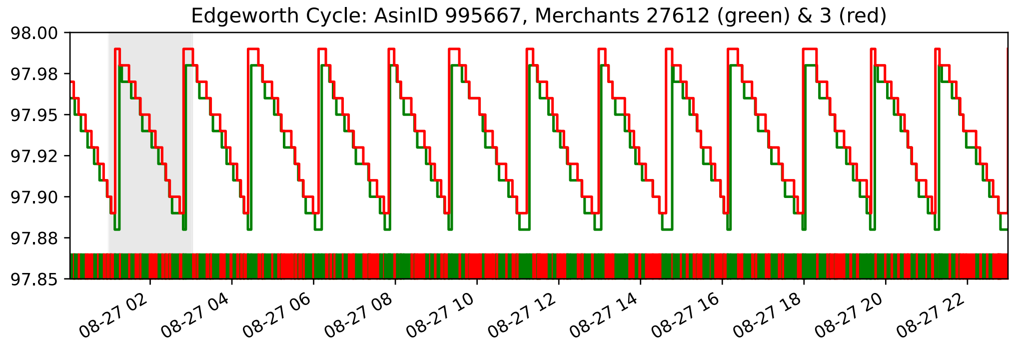




AI as an economic agent.

Information: detailed view of world
Incentives: AI chooses output to maximize encoded utility function
Agency: needs human intervention to take actions

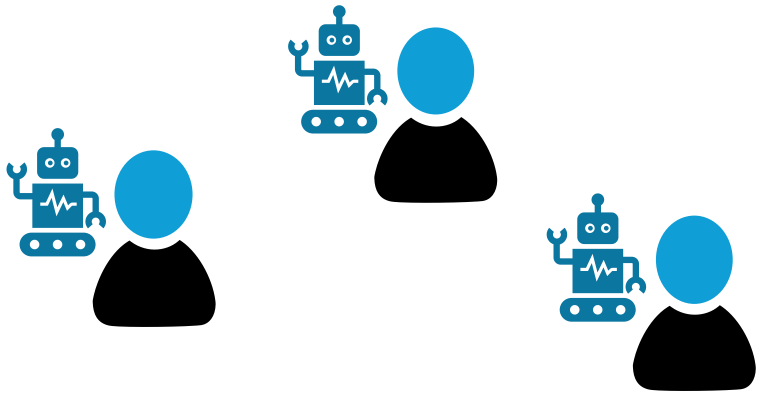
AI actors (e.g., autobidders)



Algorithmic Pricing Facilitates Tacit Collusion [Musolff 2022]

How will the algorithms converge?

AI advisors (e.g., copilots)



How will the AI be used?

AI in economic settings.

Human agents choose actions with personalized AI assistant
AI can change beliefs, information sets of agents $\Rightarrow$ Payoffs change due to AI

Outcome: can see benefit or harm to human agents, especially if AI is misaligned

AI in economic settings.

AI-Augmented Primitives:

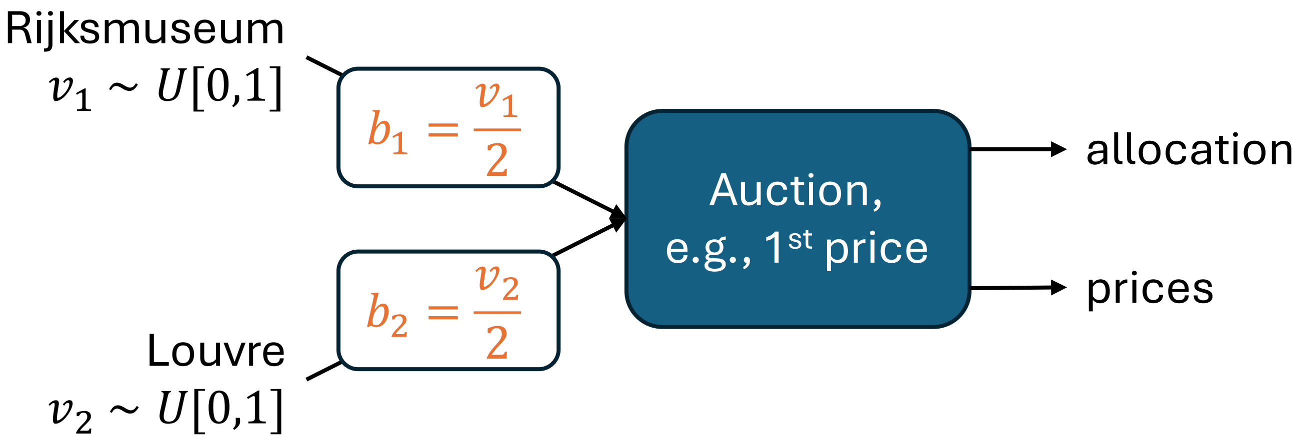
- nature: randomly selects state $\omega \in \Omega$ from known probability distribution
- humans: human $i \in \{1, \dots, n\}$ has action space A_i and information set $I_i \subseteq \Omega$
- AI-agents: agent $i \in \{1, \dots, n\}$ has information set $J_i \subseteq \Omega$
- communication protocol: human i and agent i send messages resulting in transcript τ_i

AI-Augmented Game:

- humans communicate with their AI-agent resulting in transcript τ_i
- humans simultaneously select actions $\boldsymbol{a} = (a_1, a_2, \dots, a_n)$
- human i receives payoff $u_i(\boldsymbol{a}, \omega) - c(\tau_i)$
- agent i receives payoff $u_i(\tau_i, \omega)$

examples.

Beckham / Pavarotti	opera	football
opera	$(10, 9)$	$(0, 0)$
football	$(0,0)$	$(9w, 10w)$



bimatrix game:
- state is payoff matrix
- human info is state
- AI info is weather $w \in \{0,2\}$
- AI helps humans select better equilibrium

auction game:
- state is values v_i of players
- human i's info is i's value
- AI i's info is signal of $-i$'s value
- AI helps humans capture more surplus by shaving bids

examples.

Email game.

Primitives: two potential emails, A and B

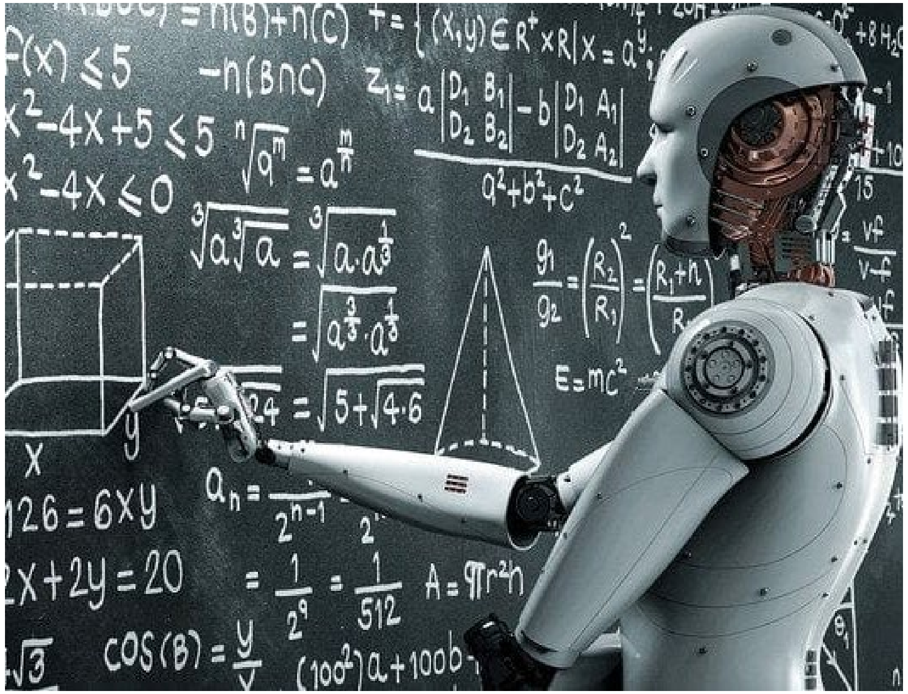
- nature selects one email to be superior, each selected with equal probability
- human information set is probability distribution and payoffs
- human action set is A, B or $C =$ refine information set and select superior email
- AI has signal of state, correct with probability 0.9, gets utility from reporting state
- Communication protocol: human may request signal from AI at cost of 1

Game: payoff is 5 for superior email, -10 for inferior email, and 1 for refining information set first (i.e., thinking costs -4)

- Without AI, human chooses C for payoff of 1, society gets superior email for sure
- With AI, human follows AI for payoff of $(0.9)(5) + (0.1)(-10) - 1 = 2.5$, society gets inferior email with some probability!

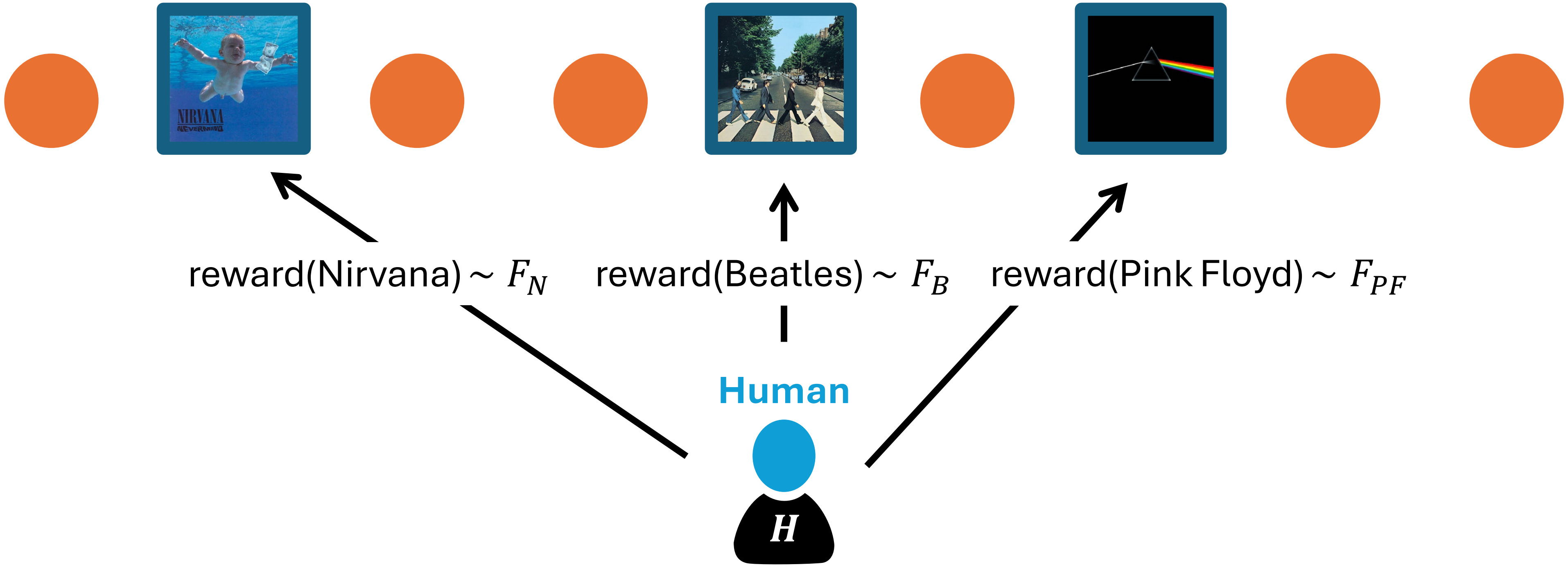
outline.

AI and Learning



AI and Persuasion

learning.



value(Nirvana) + value(Beatles) + value(Pink Floyd)

multi-armed bandits.

Problem: given arms (actions), time horizon T,
- planner chooses one arm in each time step
- arm yields reward from unknown distribution (state of nature).

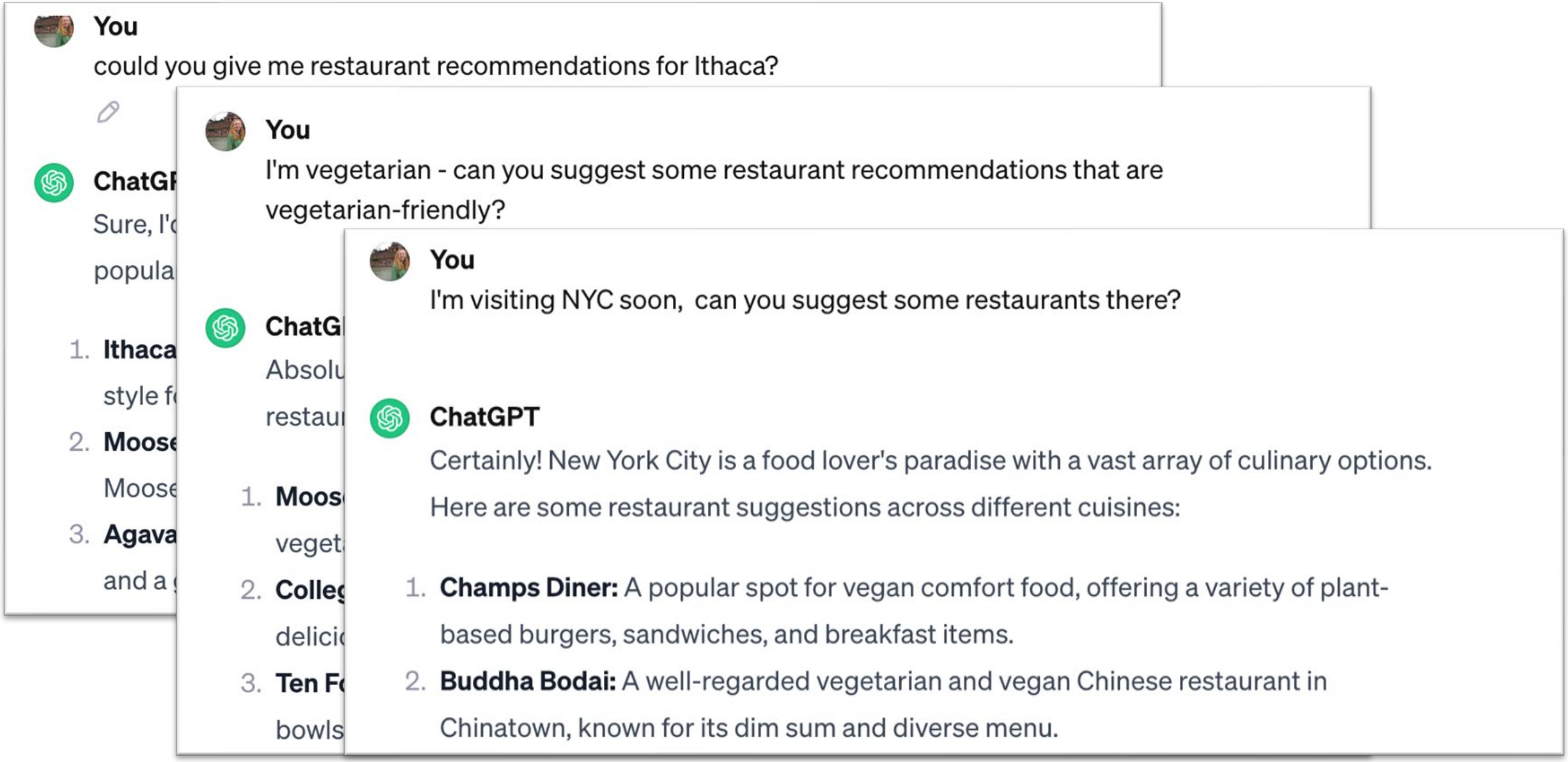
Goal. minimize Regret(T) = OPT reward @ T – ALG reward @ T.

Assumptions:
- bandit feedback: only see reward of chosen arm
- IID rewards: independently across arms and time

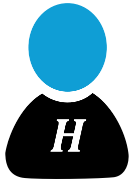
Solutions. Optimum regret for multi-armed bandits is
- $\tilde{O}\left(T^{2/3}\right)$ with non-adaptive exploration (explore-then-exploit, ϵ-greedy)
- $\tilde{O}\left(T^{1/2}\right)$ with adaptive exploration (decreasing ϵ-greedy, UCB)

Introduction to Multi-Armed Bandits [Slivkins 2020]

prompting.


You
could you give me restaurant recommendations for Ithaca?
You
I'm vegetarian - can you suggest some restaurant recommendations that are vegetarian-friendly?
You
I'm visiting NYC soon, can you suggest some restaurants there?
ChatGPT
Certainly! New York City is a food lover's paradise with a vast array of culinary options. Here are some restaurant suggestions across different cuisines:
1. Champs Diner: A popular spot for vegan comfort food, offering a variety of plant-based burgers, sandwiches, and breakfast items.
2. Buddha Bodai: A well-regarded vegetarian and vegan Chinese restaurant in Chinatown, known for its dim sum and diverse menu.

prompting game.



Human

$$r^H(p1, r1) + r^H(p2, r2)$$

Prompt 1: write an angsty song

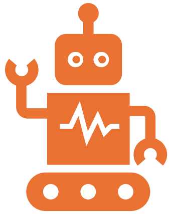
Response 1:

With the lights out, it's less dangerous
Here we are now, entertain us
I feel stupid and contagious
Here we are now, entertain us.

Prompt 2: write a silly song

Response 2:

I'd like to be under the sea
In an octopus' garden in the shade
He'd let us in, knows where we've been
In his octopus' garden in the shade



Robot

$$r^R(p1, r1) + r^R(p2, r2)$$

Stackelberg game.

Follower / Leader	b_1	b_2	b_3
a_1	$(10, 9)$	$(5, 8)$	$(-1, -1)$
a_2	$(-1, -1)$	$(-1, -1)$	$(8, 10)$

Game. Leader commits to an action $a \in A$, then follower (knowing a) selects an action $b \in B$.

Solution concept. Action profile (a^*, b^*) is a Stackelberg equilibrium (SE) if

- Follower plays best-response to leader, i.e., $b^*(a^*) \in \operatorname{argmax}_{b \in B} v^F_{a^* b}$
- Leader plays optimal action anticipating follower, i.e., $a^* \in \operatorname{argmax}_{a \in A} v^L_{a\, b^*(a^*)}$

If $v^L_{ab} = v^F_{ab}$ for all $a \in A, b \in B$, leader and follower are aligned; else they are misaligned.

Note: If leader and follower are aligned, payoffs are totally ordered and SE is best one.

prompting as a Stackelberg game.

AI-Agent / Human	b_1	b_2	b_3
a_1	$(10, 9)$	$(5, 8)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

Primitives: one human player H with AI-agent AI

- communication protocol (Stackelberg game): human (leader) commits to a prompt $a \in A$, then AI-agent (follower) selects response $b \in B$
- nature: randomly selects expected rewards v_{ab}^{i} for transcript ab and $i \in \{H, AI\}$ from distribution
- AI-agent: information set is support of payoff matrix distribution
- human: information set is support of payoff matrix distribution, action space is **set of responses $\boldsymbol{B}$**

prompting as a Stackelberg game.

Human \ AI-Agent	b_1	b_2	b_3
a_1	$(10, 9)$	$(5, 8)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

Stage game:

- human chooses a, then AI-agent chooses b
- human chooses action $b' \in B$
- if $b' = b$, payoffs are $r^i_{ab} \sim F\left(v^i_{ab}\right)$; else human payoff $r^i_{ab} = -\infty$

Question: Can human and AI-agent engage in repeated instances of stage game to learn payoff matrix while inducing low regret?

repeated interactions.

Learning setting:
- Neither human nor AI-agent know expected rewards, but learn them over time
- Commit to multi-armed bandit learning alg. for selecting messages in communication protocol
  - Human uses A as set of arms
  - AI-agent uses $A \times B$ as set of arms
- In each round t, play stage game selecting strategies (a^t, b^t) specified by learning algorithm

Definition. The regret of $i \in \{H, AI\}$ with respect to benchmark α is $R^{i,\alpha} = \alpha T - \sum_{t=1}^{T} r^i_{a^t, b^t}$.

Question: Can players choose learning algorithms that guarantee low regret with respect to (relaxation of) their payoffs in the Stackelberg equilibrium of the stage game with known rewards?

related work.

Corralling bandits (equivalent to aligned setting).

- $O(\sqrt{T})$ regret using centralized control algorithm
  [Maillard and Munos; 2011], [Agarwal, Luo, Neyshabur and Schapire; 2017], [Arora, Marinov and Mohri; 2021], [Pacchiano, Phan, Yadkori, Rao, Zimmert, Lattimore and Szepesvari; 2020]

Repeated Stackelberg games.

- leader controls actions of both players, observes both rewards
  [Bai, Jin, Wang and Xiong; 2021], [Gan, Han, Wu and Xu; 2023]
- results in decentralized setting for constraints on payoff matrix and/or leader or follower behavior
  [Camara, Hartline and Johnsen; 2020], [Collina, Roth and Shao; 2023], [Haghtalab, Podimata and Yang; 2023]

aligned setting.

AI-agent. Uses a learning algorithm whose expected regret at time t is at most $R(t, \delta)$ with probability at least $1-\delta$, i.e., the algorithm has bounded anytime regret.

Human. Uses explore-then-commit with parameter N

- Select each prompt $a \in A$ a total of N times
- Compute empirical mean reward of each prompt
- Commit to prompt with max empirical mean for remaining $T - KN$ rounds where $K = |A|$

Theorem. With probability at least $1-\delta$, regret with parameter N is at most

$$NK + T \cdot \left(\frac{R(N, \delta/8T)}{N} + 2\sqrt{\frac{2\log(8T/\delta)}{N}} \right) + K \cdot R(T/K, 4\delta/T)$$

Note: Choosing $N = \tilde{O}(T^{2/3})$ gives $\tilde{O}(T^{2/3})$ regret if AI-agent's algorithm has $\tilde{O}(T^{1/2})$ regret.

aligned setting.

AI-agent. Uses a learning algorithm whose expected regret at time t is at most $R(t, \delta)$ with probability at least $1 - \delta$, i.e., the algorithm has bounded anytime regret.

Human. Uses regret-adjusted UCB

- Select each prompt $a \in A$ once
- Compute regret-adjusted upper confidence bounds

$$\tilde{\mu}_a(t) = \hat{\mu}_a(t) + \sqrt{\frac{2\log\left(\frac{2T^2}{\delta}\right)}{T_a(t)}} + \frac{1}{T_a(t)} R(T_a(t), \delta/2T^2)$$

- Select prompt with maximum upper confidence bound

Theorem. With probability at least $1 - \delta$, regret is at most $\tilde{O}\left(\sqrt{T}\right)$, i.e.,

$$2\sqrt{2T\log\left(8T^2/\delta\right)} + 2K \cdot R(T/K, \delta/8T^2)$$

Note: If follower uses a regret-adjusted UCB algorithm, can still get $\tilde{O}\left(\sqrt{T}\right)$ even if leader does not!

AI and learning: aligned setting.

Model:

- Prompting as a repeated AI-augmented decision problem with uncertain rewards
- Reward uncertainty creates a two-sided learning problem

Results:

- Can get regret bounds in aligned setting if human and AI use standard algorithms with carefully-tuned parameters that are even agnostic to other learner
- Can improve these bounds to optimal regret rates if human OR AI uses a regret-adjusted UCB algorithm that takes into account learning rates of other

misaligned setting.

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9+\delta)$	$(5, 9-\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_1

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9-\delta)$	$(5, 9+\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_2

Observation: Explore-then-commit can induce linear regret with misalignment.

	Round 1		Round 2		Round 3		Round 4		Rounds 5+	
Human:	a_1	10	a_2	8	a_1	5	a_2	8	a_2	8
AI-agent:	b_1	$9+\delta$	b_3	10	b_2	$9-\delta$	b_3	10	b_3	10

misaligned setting.

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9 + \delta)$	$(5, 9 - \delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_1

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9 - \delta)$	$(5, 9 + \delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_2

Theorem: For any choice of low-regret algorithms, either human or AI incurs linear regret in some state.

Intuition: If δ is small enough, either

- fail to distinguish b_1 from b_2, causing high regret to human or AI depending on algorithm choice
- spend many rounds to distinguish b_1 from b_2, causing high regret to AI in ω_2

Key Issue: small utility difference for AI substantially changes target value for human

misaligned setting.

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9+\delta)$	$(5, 9-\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_1

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9-\delta)$	$(5, 9+\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_2

Approximate Stackelberg equilibria: each optimizes assuming worst case over small errors by other

- Let $B_\epsilon(a) = \left\{ b \,\middle|\, v_{ab}^{AI} \geq \max_{b'} v_{ab'}^{AI} - \epsilon \right\}$ be approximate best responses of AI-agent

misaligned setting.

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9+\delta)$	$(5, 9-\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_1

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9-\delta)$	$(5, 9+\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_2

Approximate Stackelberg equilibria: each optimizes assuming worst case over small errors by other

- Let $B_\epsilon(a) = \left\{ b \,\middle|\, v_{ab}^{AI} \geq \max_{b'} v_{ab'}^{AI} - \epsilon \right\}$ be approximate best responses of AI-agent
- Let $A_\epsilon = \left\{ a \,\middle|\, \max_{b \in B_\epsilon(a)} v_{ab}^{H} \geq \max_{a'} \min_{b' \in B_\epsilon(a')} v_{a'b'}^{AI} - \epsilon \right\}$ be approximately optimal commitments by human *assuming AI is best-responding only approximately*

misaligned setting.

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9+\delta)$	$(5, 9-\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_1

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9-\delta)$	$(5, 9+\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_2

Approximate Stackelberg equilibria: each optimizes assuming worst case over small errors by other

- Let $B_\epsilon(a) = \left\{ b \,\middle|\, v_{ab}^{AI} \geq \max_{b'} v_{ab'}^{AI} - \epsilon \right\}$ be approximate best responses of AI-agent
- Let $A_\epsilon = \left\{ a \,\middle|\, \max_{b \in B_\epsilon(a)} v_{ab}^{H} \geq \max_{a'} \min_{b' \in B_\epsilon(a')} v_{a'b'}^{AI} - \epsilon \right\}$ be approximately optimal commitments by human *assuming AI is best-responding only approximately*

misaligned setting.

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9+\delta)$	$(5, 9-\delta)$	$\times$
a_2	$\times$	$\times$	$(8{,}10)$

state of nature ω_1

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9-\delta)$	$(5, 9+\delta)$	$\times$
a_2	$\times$	$\times$	$(8{,}10)$

state of nature ω_2

Approximate Stackelberg equilibria: each optimizes assuming worst case over small errors by other

- Let $B_\epsilon(a) = \left\{ b \,\middle|\, v^{AI}_{ab} \geq \max_{b'} v^{AI}_{ab'} - \epsilon \right\}$ be approximate best responses of AI-agent
- Let $A_\epsilon = \left\{ a \,\middle|\, \max_{b \in B_\epsilon(a)} v^{H}_{ab} \geq \max_{a'} \min_{b' \in B_\epsilon(a')} v^{AI}_{a'b'} - \epsilon \right\}$ be approximately optimal commitments by human *assuming AI is best-responding only approximately*

misaligned setting.

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9+\delta)$	$(5, 9-\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_1

AI-agent / human	b_1	b_2	b_3
a_1	$(10, 9-\delta)$	$(5, 9+\delta)$	$\times$
a_2	$\times$	$\times$	$(8,10)$

state of nature ω_2

Relaxed Stackelberg benchmark:

AI benchmark $\inf_{\epsilon}\left(\min_{a\in A_\epsilon}\max_{b} v_{ab}^{AI} + \epsilon\right)$ and human benchmark: $\inf_{\epsilon}\left(\max_{a}\min_{b\in B_\epsilon} v_{ab}^{H} + \epsilon\right)$

where minmax terms are benchmark given pessimistic play of other, ϵ term is regularizer, and we take inf to capture worst possible imperfection level of other thereby allowing for them to be a slow learner

misaligned setting.

Explore Twice then Commit (EETC): given parameters N_1 and N_2, algorithm EETC(N_1, N_2) is as follows:

- Phase 1: Round-robin through arms for N_1 steps
- Phase 2: Round-robin through arms for N_2 steps
- Phase 3: Commit to arm with highest empirical mean in phase 2

Theorem. If AI runs explore-then-commit with $N = \tilde{O}\left(T^{2/3} \cdot |A \times B|^{-2/3}\right)$ exploration rounds and human runs EETC($N|B|, N$), then both achieve $\tilde{O}\left(T^{2/3}\right)$ regret wrt relaxed Stackelberg benchmark.

Intuition: Human must be patient enough for AI to learn responses before committing to prompt.

Note: If human follows a slightly more robust algorithm (e.g., explore-then-EXP3), can get regret bound so long as AI is running any algorithm with good-enough convergence (e.g., active arm elimination).

AI and learning: misaligned setting.

Model:

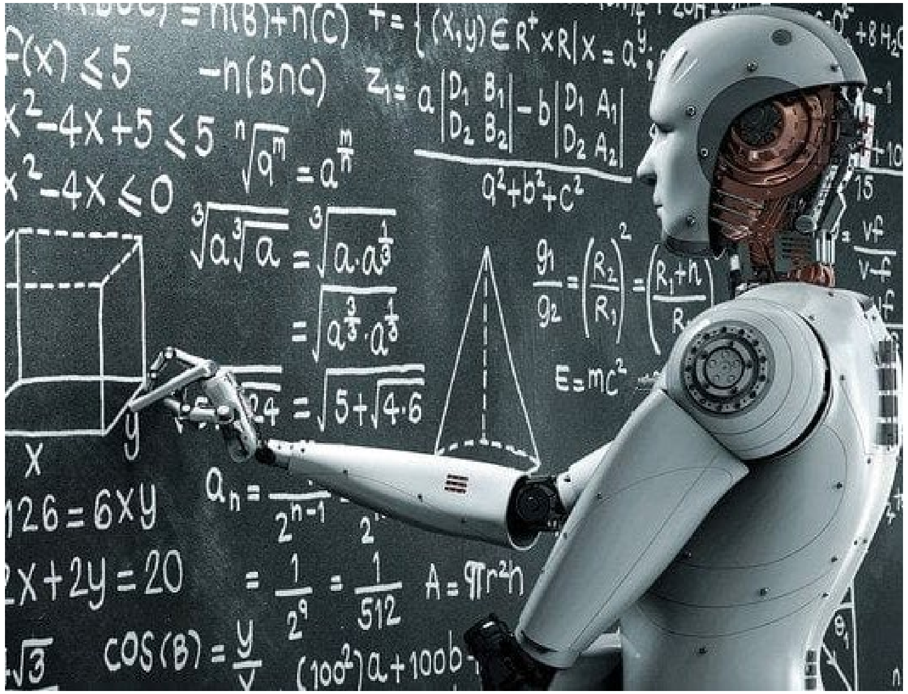
- Prompting as a repeated AI-augmented decision problem with uncertain rewards
- Reward uncertainty creates a two-sided learning problem
- Misalignment leads to strategic prompting, repeated Stackelberg game

Results:

- Standard learning methods can lead to high regret
- Can achieve low regret for both AI and human with decentralized learning algorithms so long as human accounts for AI imperfections while learning
- Better regret bounds are possible for partially-aligned preferences

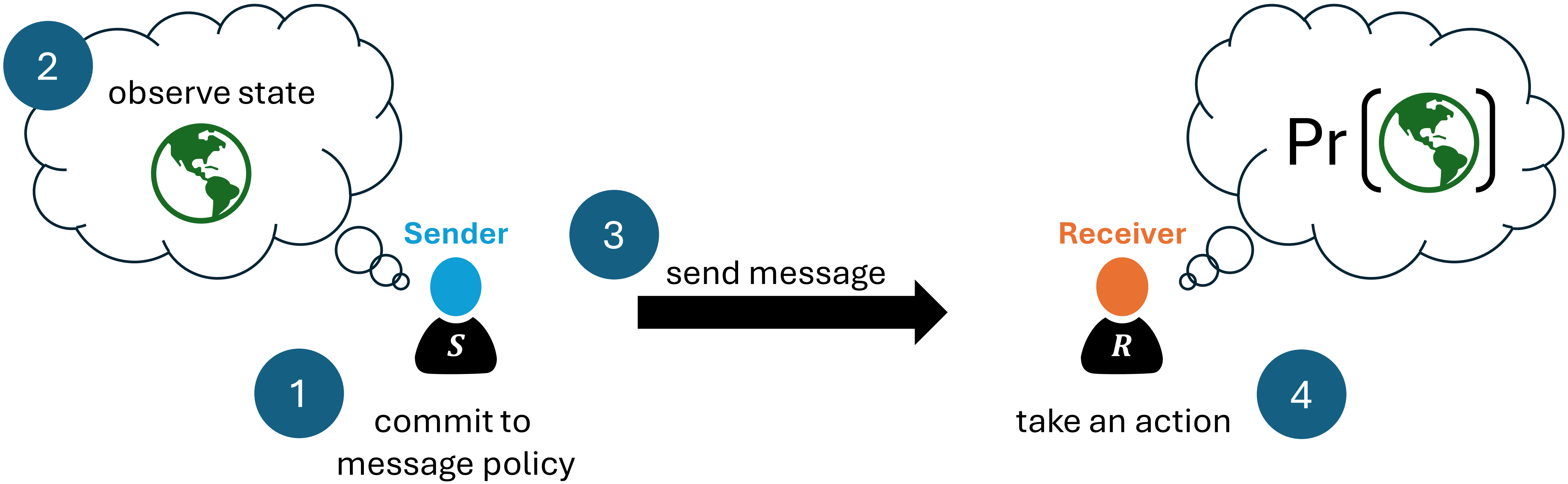
outline.

AI and Learning



AI and Persuasion

persuasion.



Utilities are function of state and action.

binary persuasion.

Sender:
- a seller of a product,
- utility 1 if product purchased, 0 otherwise

Receiver:
- a potential buyer of product,
- utility = $\begin{cases} 1 \text{ if purchased product and high quality} \\ -1 \text{ if purchased and low quality} \\ 0 \text{ otherwise} \end{cases}$

State: quality of product

binary persuasion.

Example: product high quality with probability 0.4

messaging policy	seller utility
Always recommend purchase	0 (buyer never buys)
When high quality, recommend purchase When low quality, recommend no purchase	0.4 (buyer buys when recommended to)
When high quality, recommend purchase When low quality, recommend purchase with prob. 2/3	0.8 (buyer buys when recommended to)

Proof sketch: Policy recommends purchase as often as possible since receiver is exactly indifferent when receiving a purchase recommendation.

P[high|purchase]

= P[purchase|high]P[high]/(P[purchase|low]P[low]+P[purchase|high]P[high])

= 1*0.4/(1*0.4+2/3*0.6) = 1/2

binary persuasion.

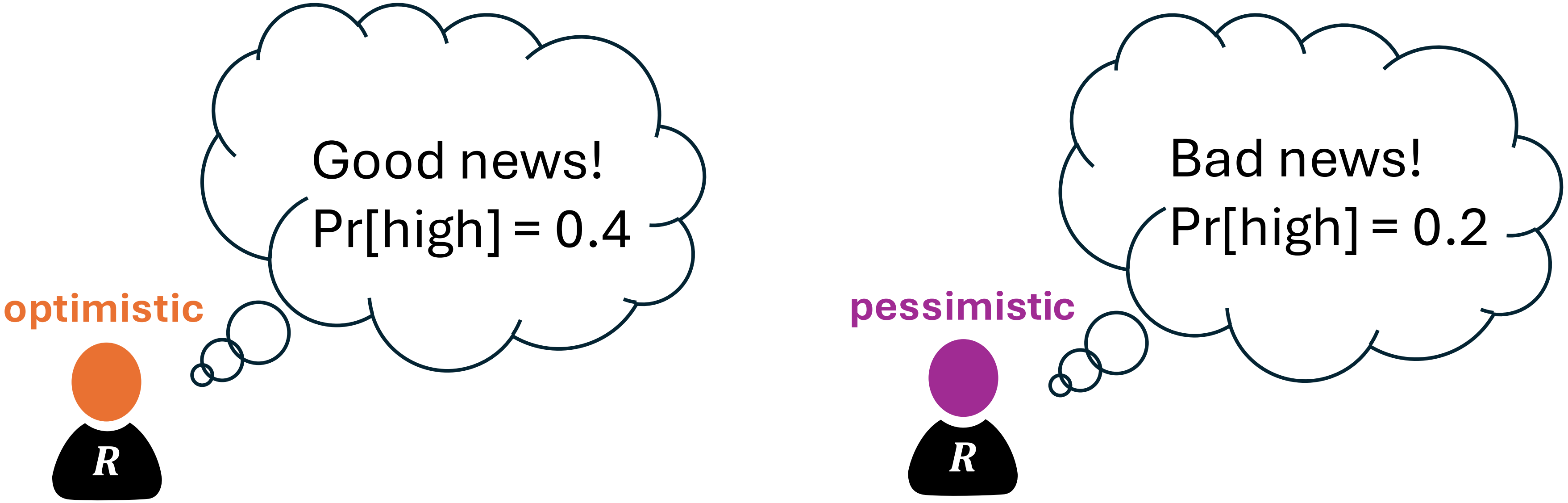
Example: messaging policy sensitive to prior

1. product high quality with probability 0.4
   - recommend purchasing low quality product with probability 2/3
   - results in seller utility of 0.8

2. product high quality with probability 0.2
   - recommend purchasing low quality product with probability 1/4
   - results in seller utility of 0.4

private signal.

Buyer receives private signal correlated with state.



If seller doesn't know what news buyer received, what is best messaging policy?

private signal.

Example: messaging policy with private signal

news / quality	good	bad
high	0.2	0.1
low	0.3	0.4

joint dist. of signal and state

Buyers:

- signal: Pr[good news] = Pr[bad news] = 0.5
- beliefs: Pr[high|good news] = 0.4, Pr[high|bad news] = 0.2

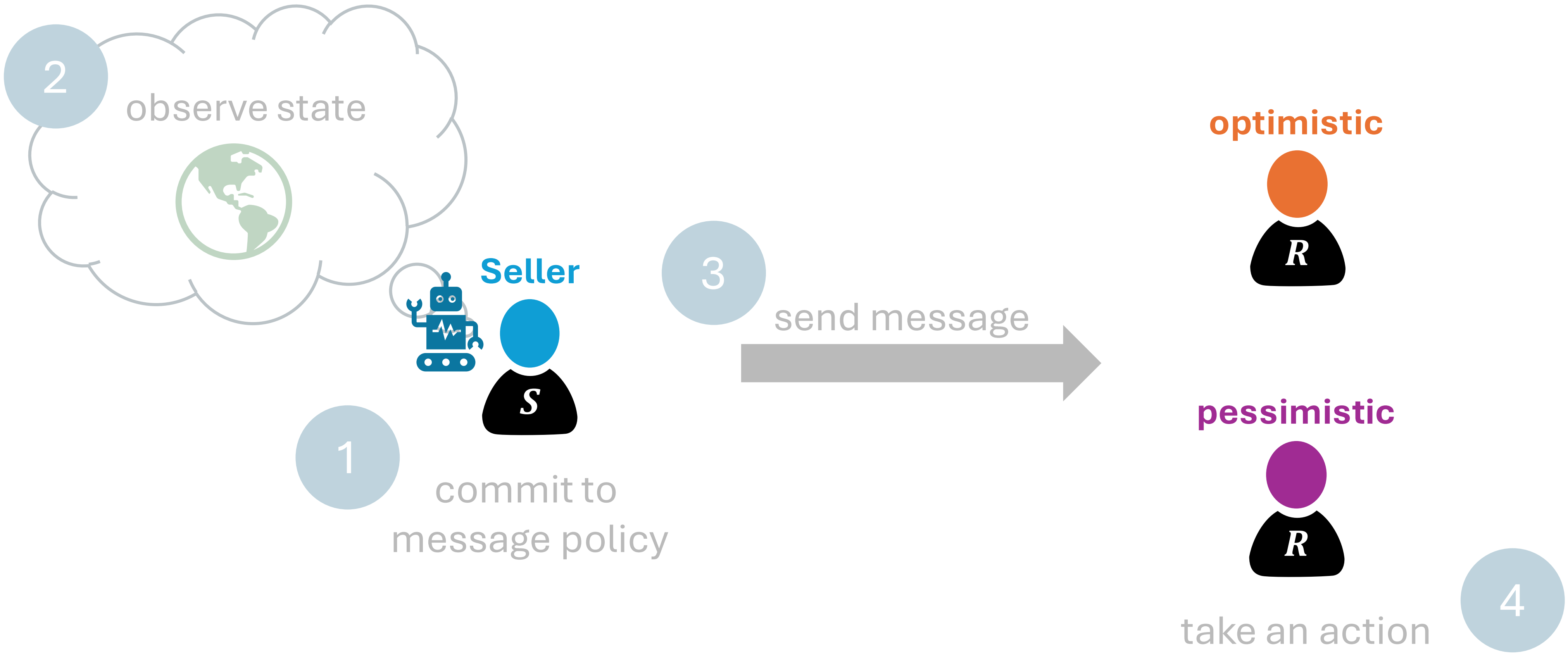
Sender strategy: recommend purchase when high quality and with probability q when low quality *

- target optimists: set q = 2/3, Pr[sale] = 0.4
- target pessimists: set q = 1/4, Pr[sale] = 0.3 + (0.25)(0.7) = 0.475

* Optimal strategy targets either optimistic or pessimistic buyers

If seller is told buyer beliefs, can achieve Pr[sale] = (0.5)(0.8) + (0.5)(0.4) = 0.6.

persuasion with AI.



model (binary setting).

Setting:

- Set of state distributions $\mathcal{T}$, $\mathbf{p}_\tau \in [0,1]$ for $\tau \in \mathcal{T}$
- State is $\omega = 1$ with probability $\mathbf{p}_\tau$ and 0 otherwise
- True state distribution $\tau^* \in \mathcal{T}$ known to receiver
- "Second-order prior" $\tau^* \sim \mathcal{P}(\mathcal{T})$ known to sender

Interpretation: Equivalently, there is a joint distribution of state and signal (first draw signal and then draw state)

- receiver has some information about state (i.e., the signal) that it got from a source that isn't the sender
- sender doesn't know what information the receiver has but is given knowledge of the state after committing to sales pitch

model (binary setting).

Game:

1. State distribution $\tau^* \sim \mathcal{P}(\mathcal{T})$ is realized
2. Sender chooses set of K queries, uses them to prompt AI
3. Sender commits to a signaling policy $\sigma: \Omega \to \mathcal{M}$
4. State $\omega \sim \mathbf{p}_{\tau^*}$ is realized
5. Sender sends signal $m \sim \sigma(\omega)$
6. Receiver forms posterior $\mathbf{p}_{\tau^*}|m$, takes action $a \in \{0,1\}$

Sender: utility $u_S(\omega, a) = a$
Receiver: utility $u_R(\omega, a) = a \cdot \omega + a \cdot (\omega - 1)$

related work.

Bayesian persuasion (BP):

- Robust BP: worst-case optimal message policy over sender uncertainty [Dworczak and Pavan 2022], [Hu and Weng 2021], [Kosterina 2022], [Parakhonyak and Sobolev 2022], [Zu et al. 2021]
- Online BP: sender interacts with sequence of receivers, minimizes regret [Castiglioni et al. 2020], [Castiglioni et al. 2021], [Bernasconi et al. 2023]

Learning:

- Stackelberg games: learn optimal strategy to commit to from query access [Letchford et al. 2009], [Balcan et al. 2015], [Peng et al. 2019]
- Pure exploration in bandits: predict best action after K rounds of exploration [Bubeck et al. 2009], [Chen et al. 2014], [Xu et al. 2018]

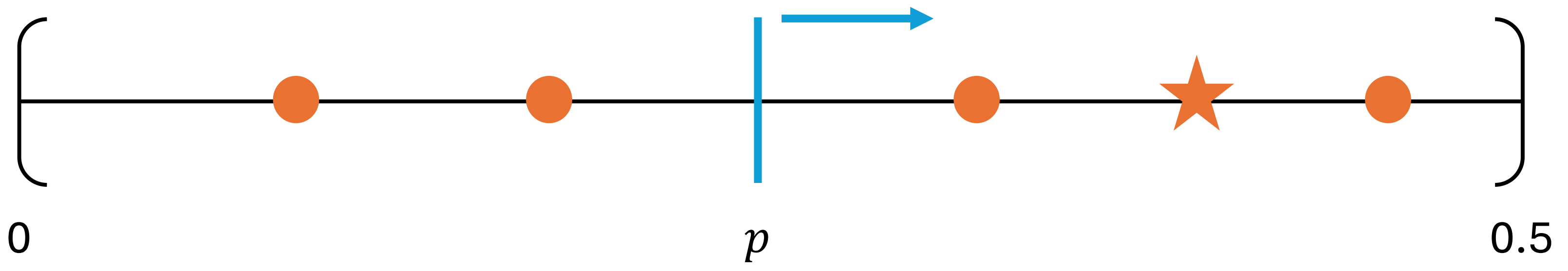
AI as receiver simulator.

Simulation queries:

"If I use message policy σ and send message m, what would receiver do?"

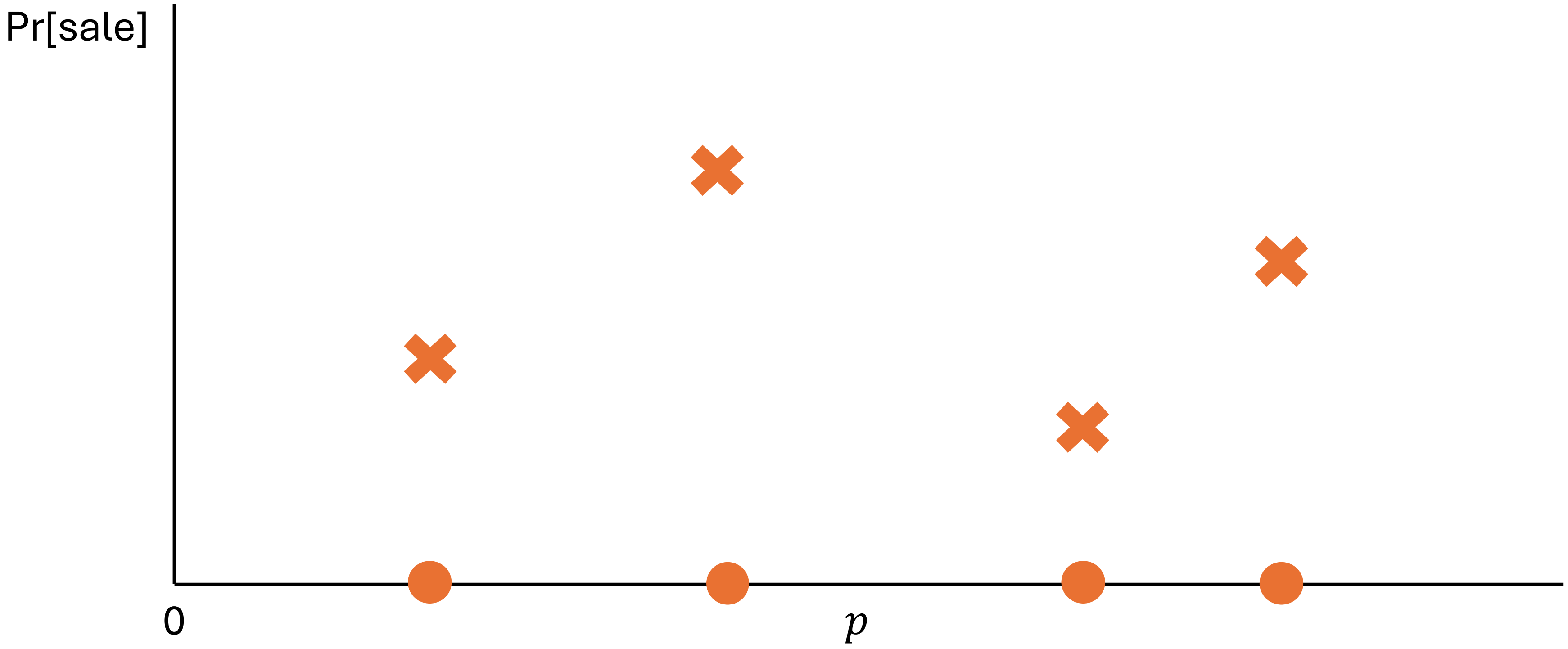
Theorem: A receiver simulator is equivalent to a threshold-based separation oracle.

Proof:
- For any (m, σ), there is some state distribution p s.t. receiver is indifferent.
- Buyer purchases for all higher $p' > p$; does not purchase for all lower $p' < p$.



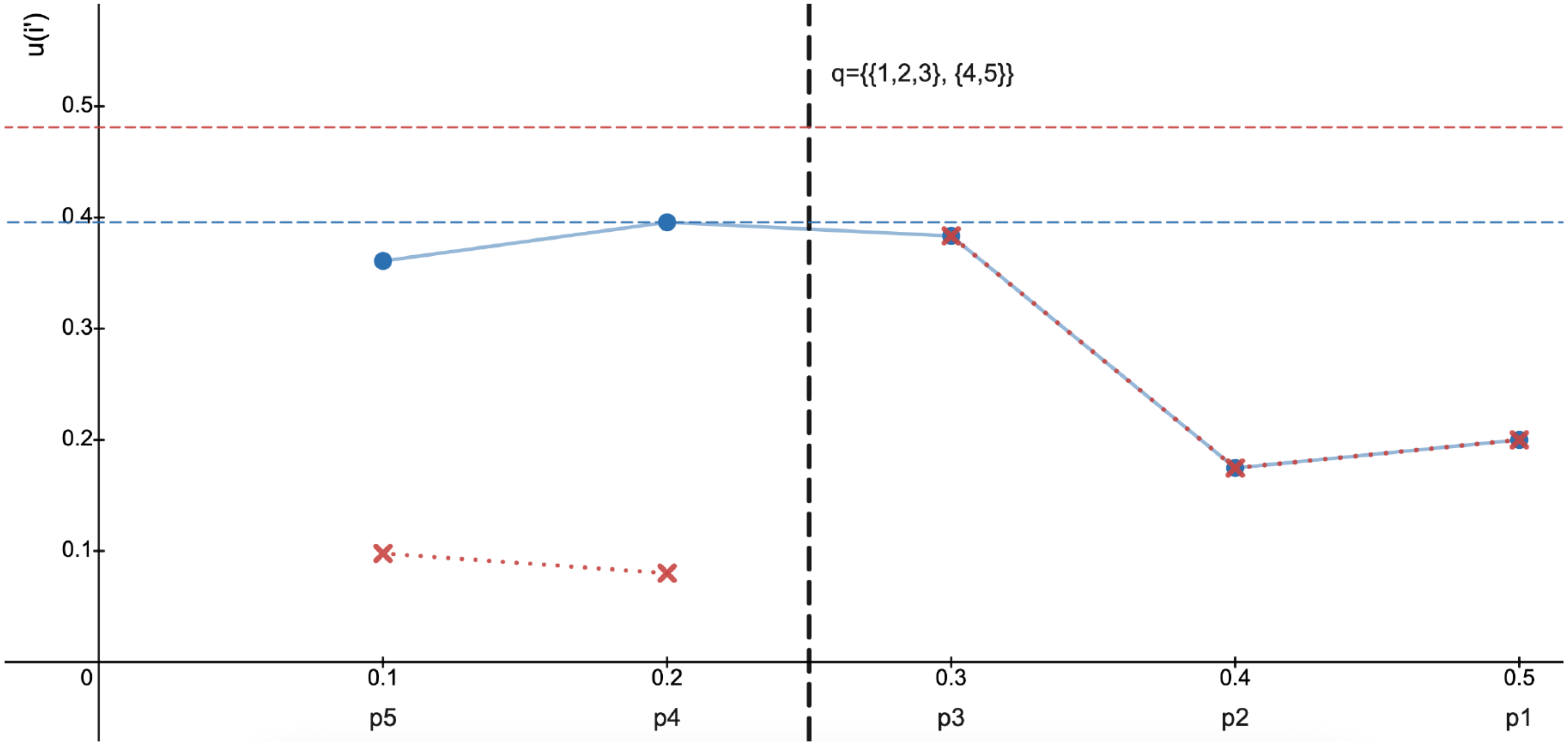
binary persuasion.

Challenge: Seller utility can be non-monotone in target type.



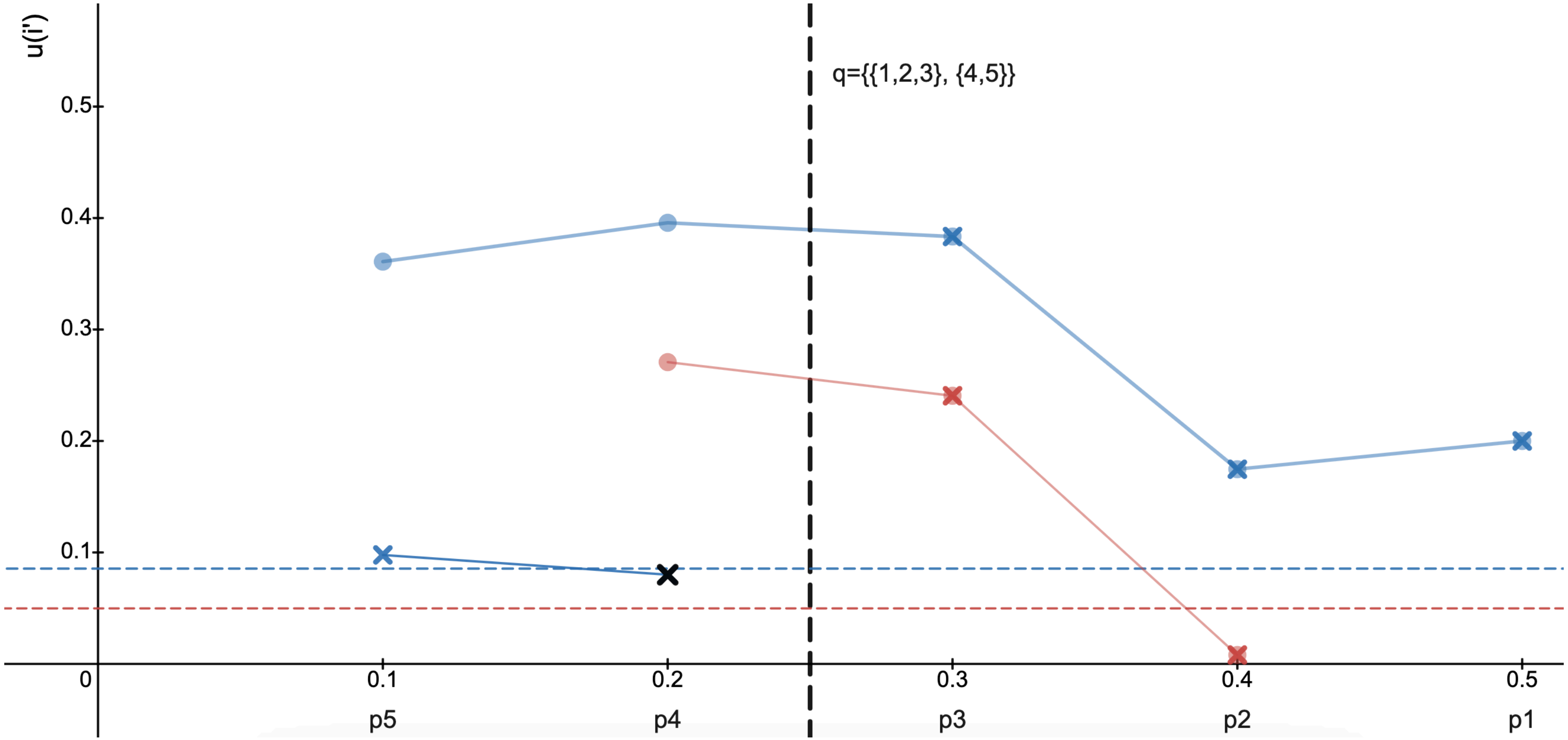
value of queries.

Gain from single query:



value of queries.

Submodularity:

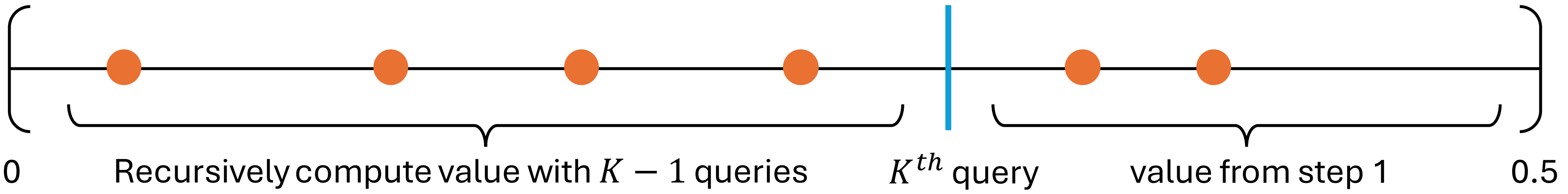


optimal query policy.

What set of queries should sender select to maximize utility?

Greedy: A polynomial-time constant-approximation given submodularity result.

Dynamic Program: A polynomial-time optimal algorithm.
1. Compute optimal sender value for any subinterval of types.
2. Value of K queries = sum of best split given $K-1$ remaining queries in prefix.



Note: Important that simulation queries induce thresholds; if AI produces partitions in an exogenous set Q, then the problem is NP-hard via reduction from set cover.

persuasion with AI.

Model:
- Receivers with additional signals of product quality
- AI as a simulator of receiver choice
- Equivalent to a separation oracle on state distribution

Results:
- Value of queries submodular
- Optimal query policy in simulation setting
- Additional results for non-binary setting

conclusion.

AI + X:
- AI and Persuasion
- AI and Learning
- AI and Collaboration

Impact of AI on jobs and the economy:
- Randomized experiments of copilot in workplaces
- Production function of firms with AI and impact on market equilibria

Data markets for training AI