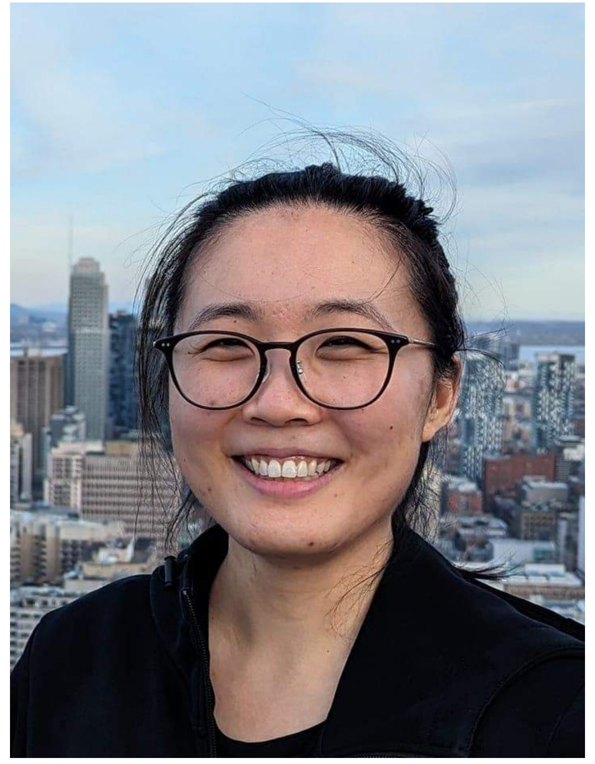
Interactive decision-making via autoregressive generation

Hong Namkoong

Columbia University



Tiffany Cai

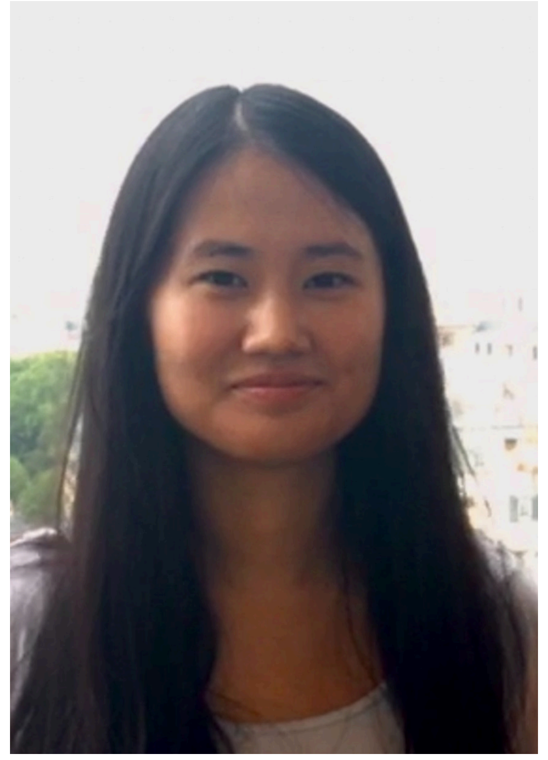
Daksh Mittal

Dan Russo

Jimmy Wang

Naimeng Ye

Rich Zemel



Kelly Zhang

Tom Zollo

Progress in AI

2014

Goodfellow et al. (2014) – Generative Adversarial Networks

2015

Radford, Metz, and Chintala (2015) – Unsupervised Representation Learning with Deep Convolutional GANs

2016

Liu and Tuzel (2016) – Coupled GANs

2022 Image generated with the prompt: *"A Pomeranian is sitting on the King's throne wearing a crown. Two tiger soldiers are standing next to the throne."*

Saharia et al. (2022) – Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding (Google's Imagen)

AI = algorithms + **data**

90% of AI research…

- Optimization algos
- Architectures
- Objective functions

First-order issue

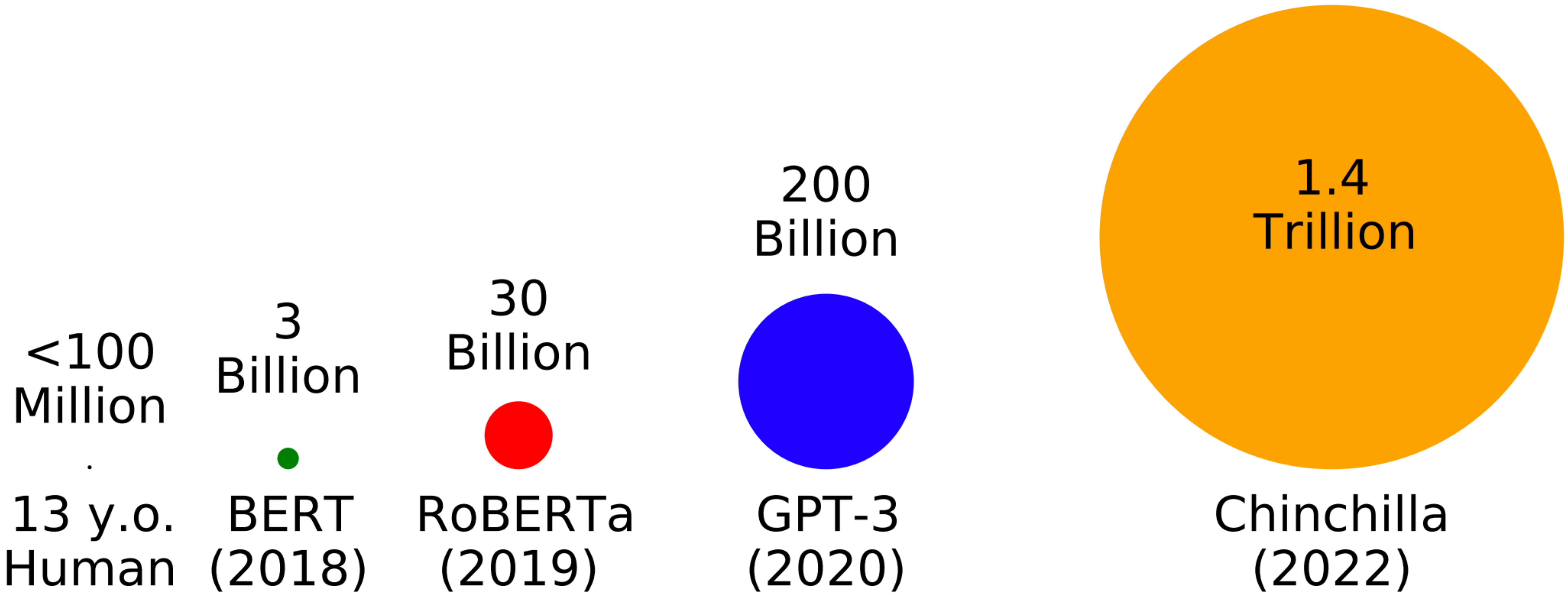
Progress in AI driven by scaling data

Language modeling

The student consistently confuses the value of digits in multi-digit numbers, struggles to regroup when subtracting across zeros, and relies heavily on counting strategies. These are classic indicators of ___________.

Possible answers: weak number sense, limited place value understanding, poor procedural fluency

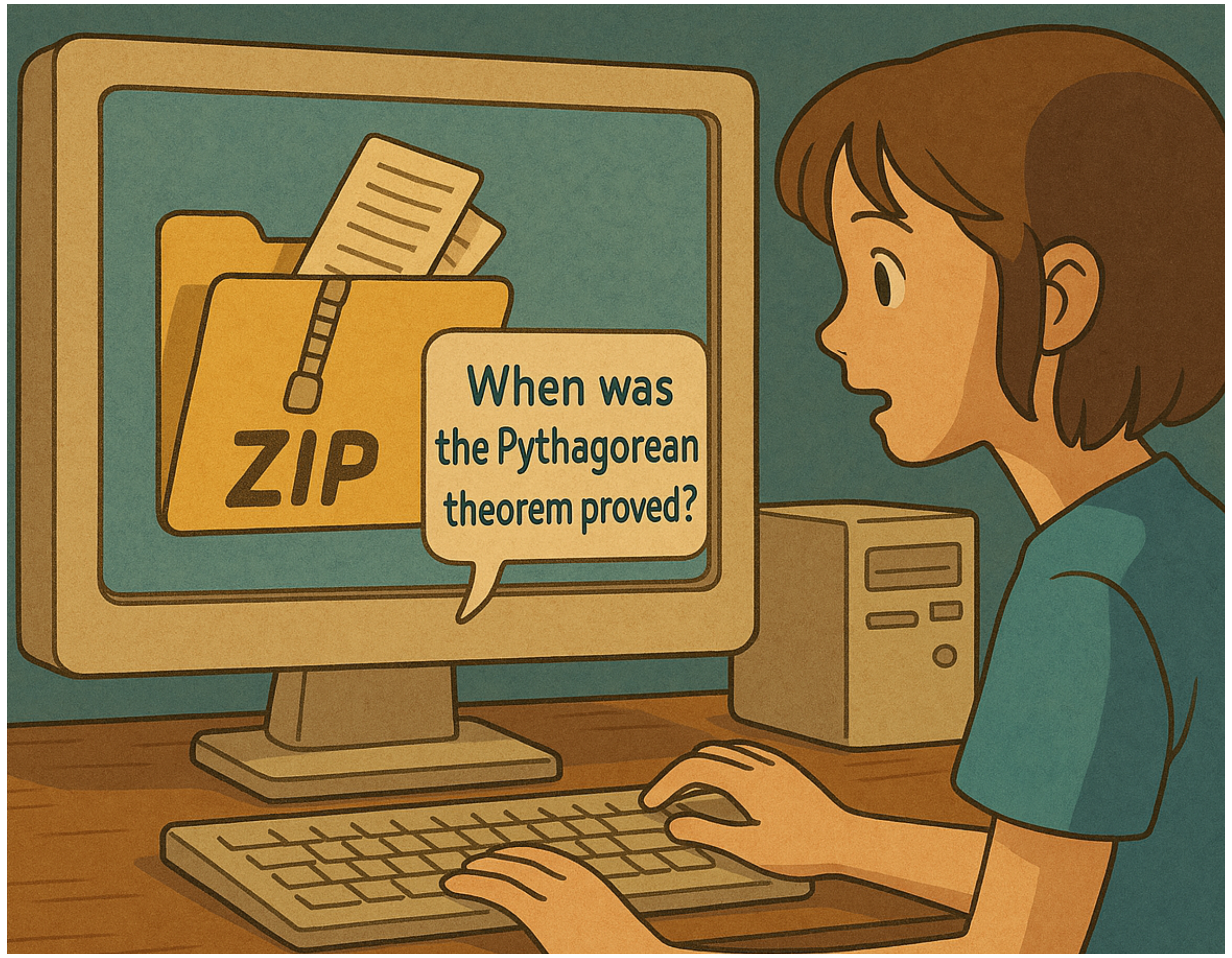
Scaling data



My view of current AI capabilities

Natural language interface is a disruptive force

- Compression engine (knowledge distillation)
  - Internet is zipped down to 1 trillion parameters
- LLMs provides a **new UX to computing**
  - Retrieve information in natural language
  - E.g., coding as translation: English => Python



Agents need to discover & improve

- I value the ability to **improve and learn,** not memorize routines

Current AI development

OpenAI collects high quality knowledge and distills it into the model

The student consistently confuses the value of digits in multi-digit numbers, struggles to regroup when subtracting across zeros, and relies heavily on counting strategies. These are classic indicators of poor procedural fluency.

Agents need to discover & improve

- I value the ability to **improve and learn,** not memorize routines

Current AI development

OpenAI collects high quality knowledge and distills it into the model

The student consistently confuses the value of digits in multi-digit numbers, struggles to regroup when subtracting across zeros, and relies heavily on counting strategies. These are classic indicators of poor procedural fluency.

Where I want to go…

A tutoring agent constantly faces new students and must gather information to quickly assess their proficiency

Ask a place value question → Notice hesitation with regrouping → Pose subtraction problem with borrowing

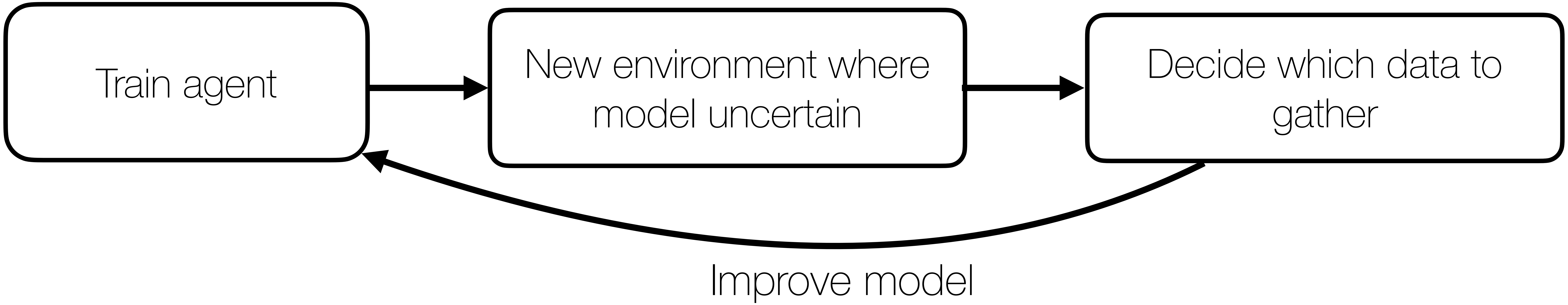
Rich Sutton

We want AI agents that can discover like we can, not which contain what we have discovered. Building in our discoveries only makes it harder to see how the discovering process can be done.

Turing Award Winner in 2025

Agents interacting with the real-world

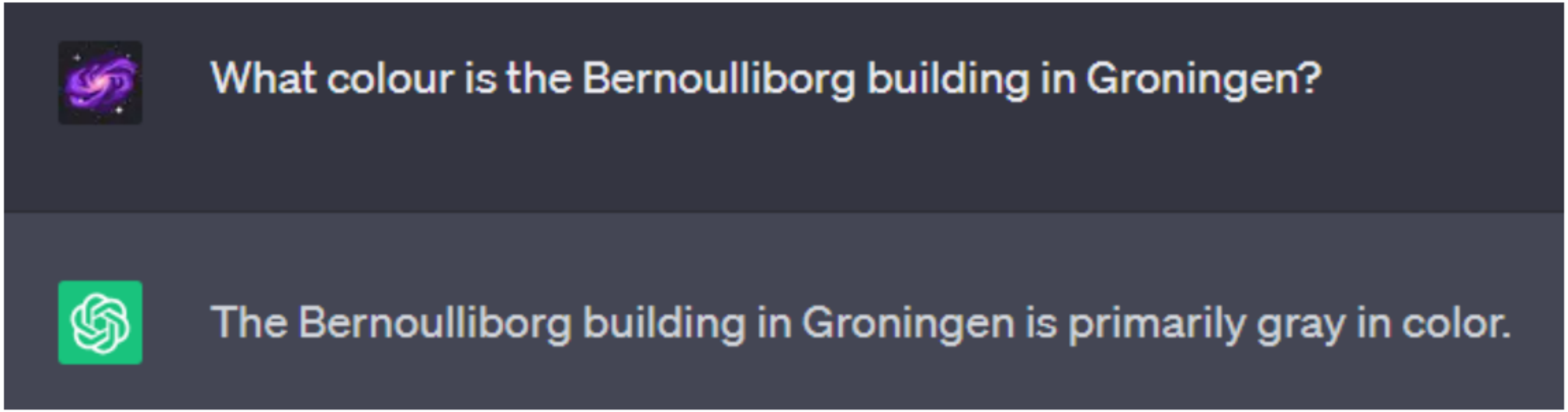
- Real-world decision-making needs to deal with a continual lack of data
  - Ever-present distribution shift from new student, patient, user, item
- AI must articulate uncertainty and act to resolve it



Uncertainty quantification

ChatGPT Prompting Cannot Estimate Predictive Uncertainty

Pelucchi and Valdenegro (2023)



ChatGPT Prompting Cannot Estimate Predictive Uncertainty

Pelucchi and Valdenegro (2023)

Introducing deep research

An agent that uses reasoning to synthesize large amounts of online information and complete multi-step research tasks for you. Available to Pro users today, Plus and Team next.

"It may struggle with distinguishing authoritative information from rumors, and currently shows weakness in confidence calibration, often failing to convey uncertainty accurately."

https://openai.com/index/introducing-deep-research/

Uncertainty quantification

- Several line of work tackle UQ
  - Bayesian neural networks, GPs, ensembles, epistemic neural nets, conformal prediction, multi-calibration…many other interesting ideas
- But these ideas have not materialized in the form of scalable models that can deal with unstructured information such as natural language

Uncertainty quantification

- Several line of work tackle UQ
  - Bayesian neural networks, GPs, ensembles, epistemic neural nets, conformal prediction, multi-calibration…many other interesting ideas
- But these ideas have not materialized in the form of scalable models that can deal with unstructured information such as natural language
- Why? They cannot scale with internet-scale datasets

AI = algorithms + data

Classical Approach

Model “environment” first, then pass onto quantity of interest

Prior: latent “environment” drawn $\theta \sim \pi(\,\cdot\,)$

Likelihood: observations generated by $\mathbb{P}(\text{ obs } | \text{ context}, \theta\ )$

- As you gather data, infer what the environment looks like

Posterior $\mathbb{P}(\ \theta \mid \text{history}\ )$

- Bayes rule provides a natural modeling language

Example: adaptive student assessment

Why probabilistic modeling is hard

- Latent parameter θ = student's "math proficiency"
- Posterior P(math proficiency | Q&A, prior info on student)

Prior information on student

Amin

Amin generally solves single-digit division problems fluently and uses repeated subtraction when working with larger numbers. When asked to explain his reasoning, he often relies on 'what feels right' rather than written strategies. He's more confident when problems are framed in real-life contexts, like sharing food or money, and shows persistence even when confused.

Prior???

Likelihood???

Example: adaptive student assessment

Why probabilistic modeling is hard

- Latent parameter θ = student's "math proficiency"
- Posterior P(math proficiency | Q&A, prior info on student)

Latent has no physical meaning!

Hard to check whether your unicorn is better than mine

Why probabilistic modeling is hard

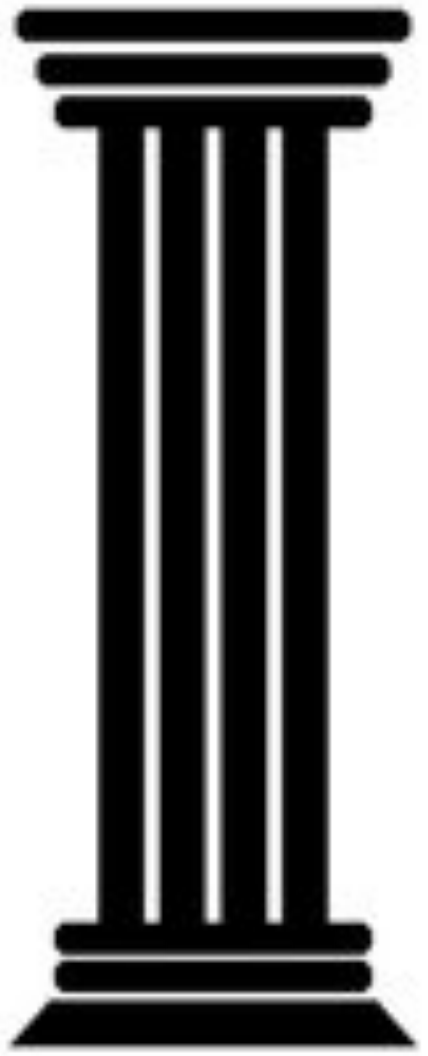
- Two pillars of ML

    1. Optimize loss on web-scale training data

    2. Test engineering innovations based on val loss

- Hard to fit aforementioned ideas into this umbrella

PyTorch



Today: Adopt these principles to quantify uncertainty

Today: A fresh yet classical perspective

Uncertainty comes from **missing data** yet to be observed

Today: A fresh yet classical perspective

Uncertainty comes from **missing data** yet to be observed

Prior information on student

Amin

Amin generally solves single-digit division problems fluently and uses repeated subtraction when working with larger numbers. When asked to explain his reasoning, he often relies on 'what feels right' rather than written strategies. He's more confident when problems are framed in real-life contexts, like sharing food or money, and shows persistence even when confused.

Today: A fresh yet classical perspective

Uncertainty comes from **missing data** yet to be observed

Prior information on student

Question: A juice bottle has 3.9 liters. If I share it equally among 13 friends how many liters does each get?

Question: If 13 aunts each give Timmy $0.75 for Christmas, how much does he have?

...

Amin generally solves single-digit division problems fluently and uses repeated subtraction when working with larger numbers. When asked to explain his reasoning, he often relies on 'what feels right' rather than written strategies. He's more confident when problems are framed in real-life contexts, like sharing food or money, and shows persistence even when confused.

Amin

Missing data as source of uncertainty

Today: A fresh yet classical perspective

Uncertainty comes from **missing data** yet to be observed

Prior information on student

Amin

Amin generally solves single-digit division problems fluently and uses repeated subtraction when working with larger numbers. When asked to explain his reasoning, he often relies on 'what feels right' rather than written strategies. He's more confident when problems are framed in real-life contexts, like sharing food or money, and shows persistence even when confused.

Question: A juice bottle has 3.9 liters. If I share it equally among 13 friends how many liters does each get?

A. 0.03 liters
B. 0.3 liters
C. 0.39 liters
D. 3 liters ✓
E. Approximately 3 liters

Question: If 13 aunts each give Timmy $0.75 for Christmas, how much does he have?

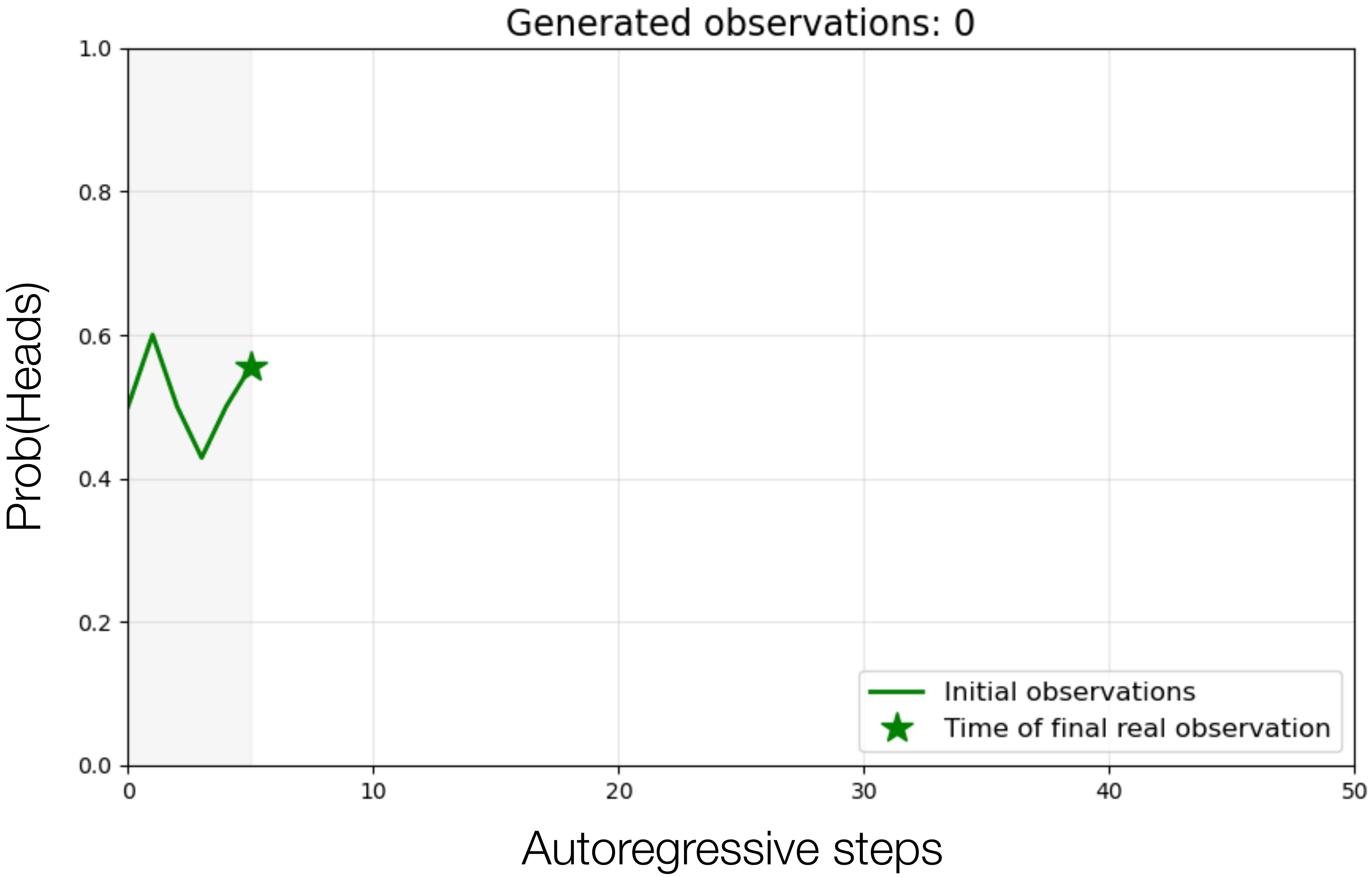
A. $1.75
B. $9.75 ✓
C. $10.75
D. $12.75
E. $13.75

...

Autoregressive generation as posterior inference

Uncertainty = variability in generated trajectories

- Conditioned on observed data, autoregressively generate imagined future data
- Compute quantity of interest
- Repeat to get histogram

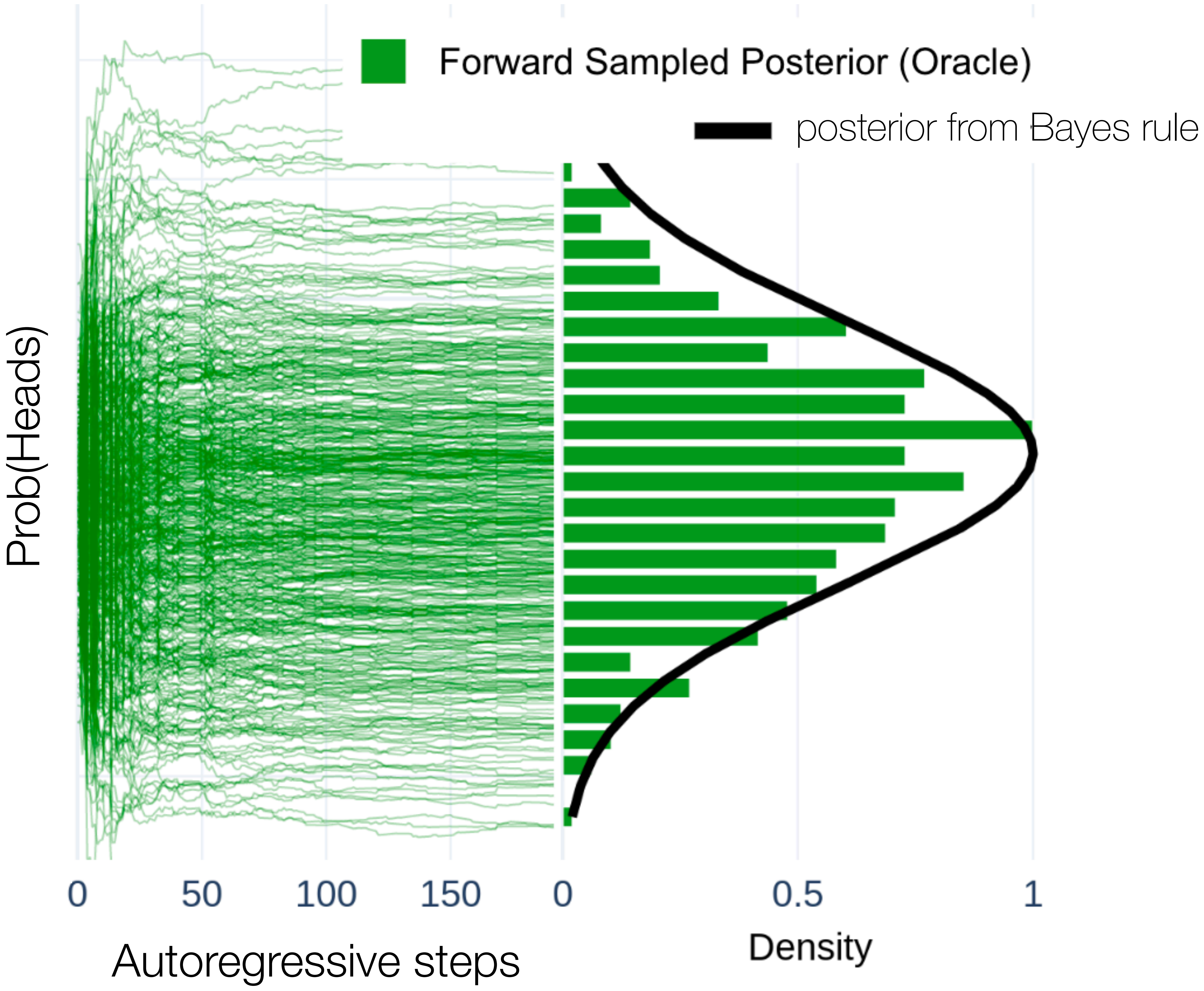


Uncertainty = variability in generated trajectories

- Conditioned on observed data, autoregressively generate imagined future data

- Compute quantity of interest

- Repeat to get histogram



Example: student assessment

Generalization to language-based problems

Amin generally solves single-digit division problems fluently and uses repeated subtraction when working with larger numbers. When asked to explain her reasoning, he often relies on ‘what feels right’ rather than written strategies. He’s confident when……

A juice bottle has 3.9 liters. If I share it equally among 13 friends how many...

I divided 3.9 by 13 and got 3, so I think each friend gets 3 liters. Answer: D

⋮

If 13 aunts each give Timmy $0.75 for Christmas, how much does he have?

I multiply 13 × 0.75 and get $9.75 total. Answer: B

Example: student assessment

Generalization to language-based problems

Amin generally solves single-digit division problems fluently and uses repeated subtraction when working with larger numbers. When asked to explain her reasoning, he often relies on 'what feels right' rather than written strategies. He's confident when……

A juice bottle has 3.9 liters. If I share it equally among 13 friends how many...

I divided 3.9 by 13 and got 3, so I think each friend gets 3 liters. Answer: D

If 13 aunts each give Timmy \$0.75 for Christmas, how much does he have?

I multiply 13×0.75 and get \$9.75 total. Answer: B

⋮

Problem bank

If Erin makes \$1375 per 11-hour work day, how much does she make per hour?

A water bottle has 0.60 liters. Pour it equally into 2 cups. How many liters per…

Example: student assessment

Generalization to language-based problems

Amin generally solves single-digit division problems fluently and uses repeated subtraction when working with larger numbers. When asked to explain her reasoning, he often relies on ‘what feels right’ rather than written strategies. He’s confident when……

A juice bottle has 3.9 liters. If I share it equally among 13 friends how many...

I divided 3.9 by 13 and got 3, so I think each friend gets 3 liters. Answer: D

⋮

If 13 aunts each give Timmy $0.75 for Christmas, how much does he have?

I multiply 13 × 0.75 and get $9.75 total. Answer: B

Autoregressively simulate from LLM

Problem bank

If Erin makes $1375 per 11-hour work day, how much does she make per hour?

I divide $1375 by 11 hours, which gives me $125 per hour. Answer: A

A water bottle has 0.60 liters. Pour it equally into 2 cups. How many liters per…

I divide 0.60 by 2 and get 0.06 liters per cup. Answer: C

⋮

Example: student assessment

Generalization to language-based problems

Amin generally solves single-digit division problems fluently and uses repeated subtraction when working with larger numbers. When asked to explain her reasoning, he often relies on ‘what feels right’ rather than written strategies. He’s confident when……

A juice bottle has 3.9 liters. If I share it equally among 13 friends how many...

I divided 3.9 by 13 and got 3, so I think each friend gets 3 liters. Answer: D

⋮

If 13 aunts each give Timmy $0.75 for Christmas, how much does he have?

I multiply 13 × 0.75 and get $9.75 total. Answer: B

Autoregressively simulate from LLM

Problem bank

If Erin makes $1375 per 11-hour work day, how much does she make per hour?

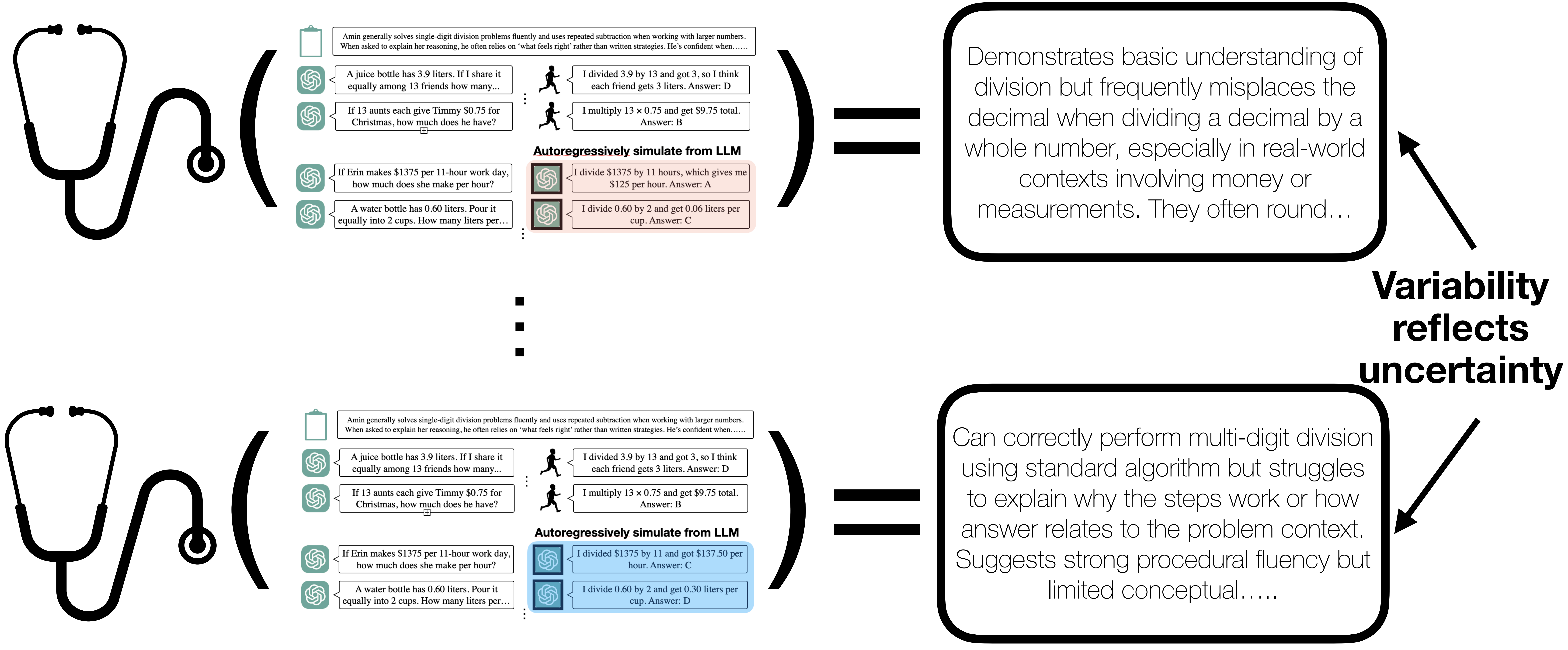
I divided $1375 by 11 and got $137.50 per hour. Answer: C

A water bottle has 0.60 liters. Pour it equally into 2 cups. How many liters per…

I divide 0.60 by 2 and get 0.30 liters per cup. Answer: D

⋮

Uncertainty = variability in inferred diagnosis



Importance of autoregressive generation

- Students make mistakes / guess when they don’t know the correct answer

answer = f(question, proficiency) + noise

- Autoregressive generation is critical for distinguishing proficiency from noise
  - Averaging over multiple Q&A washes out aleatoric noise (irreducible for each Q)
  - Remaining correlation reflect epistemic uncertainty (reducible with data)

Related work

TLDR: modeling joint distribution over observations challenging until 2018

- Geisser ('71), Rubin ('78) espouse similar idea, but found it "too burdensome to be worthwhile"; resorts to typical latent variable modeling
- Bayesians favor De Finneti's deep theorem over our simpler foundations
  - De Finnetti ('29): exchangeable sequence modeling = latent factor models
  - Line of work on predictive view [Berti et al. ('98, '21, '22), Fortini et al. ('14, '23), Fong et al. ('23)]
- Transformers-as-Bayes papers fail to isolate epistemic uncertainty [Muller et al., ('24), Nguyen & Grover ('23)]

Modeling by loss minimization

Probabilistic modeling as training a sequence model

- Modeling primitive: given past data, prob of next observation

$$-\sum_{\text{questions}} \log \widehat{p}\left(\text{answer} \mid \text{question}, \text{past Q\&A}, \text{student info}\right)$$

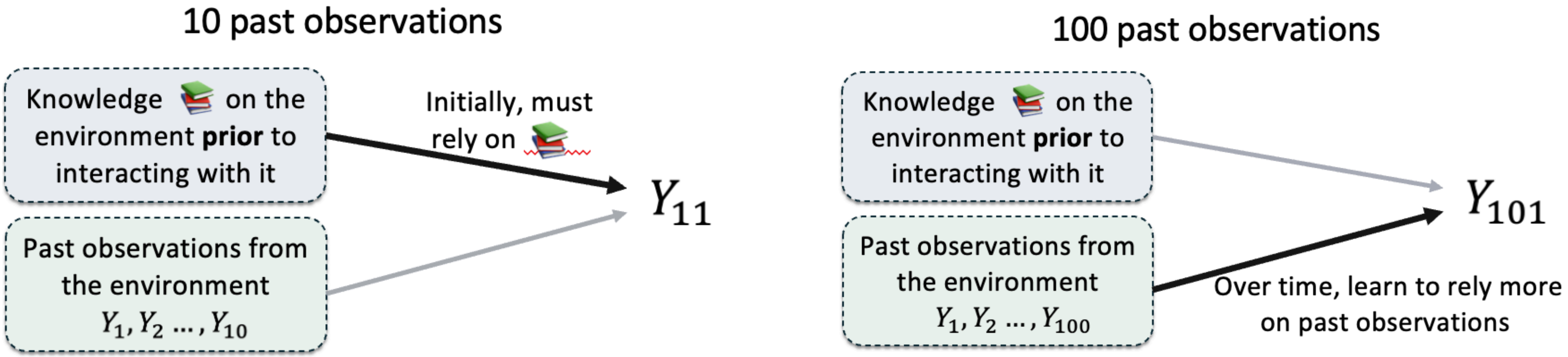
Sequence prediction loss
a.k.a. neg. log likelihood

Probabilistic modeling as training a sequence model

- Modeling primitive: given past data, prob of next observation

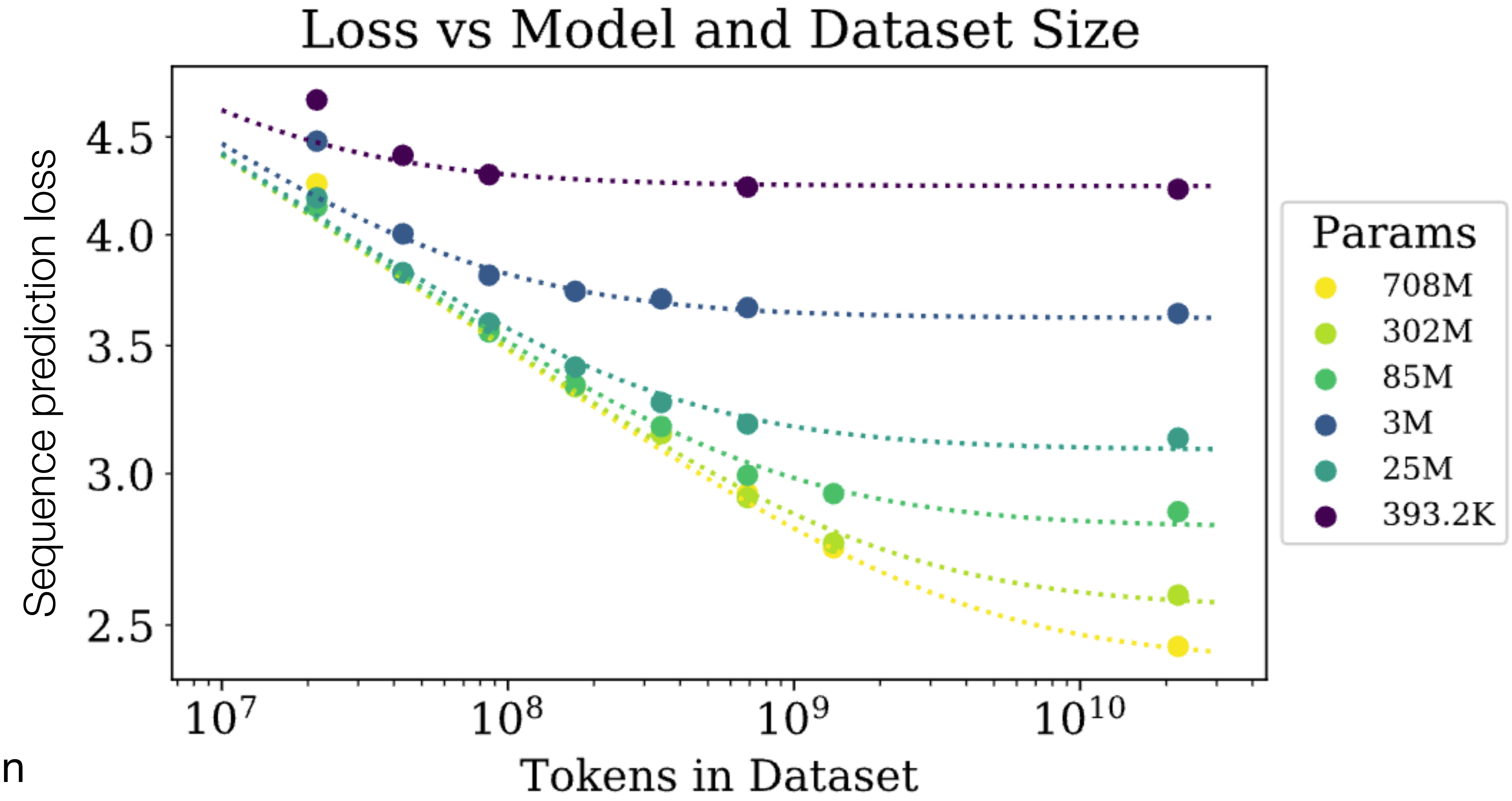
$$-\sum_{\text{questions}} \log \hat{p}\left(\text{answer} \mid \text{question, past Q\&A, student info}\right)$$

Sequence prediction loss
a.k.a. neg. log likelihood



Great news! Humanity is getting really good at training sequence models.

Scaling laws



Same loss as in previous slide

Kaplan et al. (2020), OpenAI

Scaling laws

Premise of our framework: sequence loss can be optimized offline with big data

Scaling laws

Premise of our framework: sequence loss can be optimized offline with big data

OpenAI to raise $40 billion in SoftBank-led round to boost AI efforts

By **Jaspreet Singh** and **Harshita Mary Varghese**

April 1, 2025 10:38 AM EDT · Updated a day ago

Reuters

U.S.

Three Mile Island nuclear plant to reopen, sell power to Microsoft

CBS NEWS

Updated on: September 21, 2024 / 9:54 PM EDT / CBS/AFP

Updating beliefs with more information

Traditional probabilistic modeling
Uncertainty comes from unknown parameters

Ingredients	Prior Likelihood	P(parameter) P(Y | 📚, parameter)
Model of uncertainty	P (parameter | 📚 , $Y_{1:10}$)	
Updating beliefs	Posterior inference, e.g., MCMC	

This work: predictive view
Uncertainty comes from missing future data

Ingredients	Autoregressive probabilities P(Y_t | 📚, $Y_{<t}$)
Model of uncertainty	P ($Y_{11:100}$ | 📚 , $Y_{1:10}$)
Updating beliefs	In-context learning (add to prompt)

Posterior update trivial under our framework!

Predictive loss controls posterior quality

Predictive loss controls posterior quality

$\tau^\star$

Prior information	
Questions	Answers
Q1	A1
Q2	A2
Q3	A3
Q4	A4
Q5	A5
Q6	A6
Q7	A7
Q8	A8

$\sim p^\star(\text{ data } | \text{ past })$

$\widehat{\tau}$

Prior information	
Questions	Answers
Q1	A1
Q2	A2
Q3	A3
Q4	A4
Q5	A5
Q6	A6
Q7	A7
Q8	A8

$\sim \widehat{p}(\text{ data } | \text{ past })$

Predictive loss controls posterior quality

$\tau^\star$

Prior information	
Questions	Answers
Q1	A1
Q2	A2
Q3	A3
Q4	A4
Q5	A5
Q6	A6
Q7	A7
Q8	A8

$\sim p^\star(\text{ data } \mid \text{ past })$

$\widehat{\tau}$

Prior information	
Questions	Answers
Q1	A1
Q2	A2
Q3	A3
Q4	A4
Q5	A5
Q6	A6
Q7	A7
Q8	A8

$\sim \widehat{p}(\text{ data } \mid \text{ past })$

Questions drawn independently from problem bank; can be non-stationary

Predictive loss controls posterior quality

Sequence prediction loss

a.k.a. neg. log likelihood

$$\ell(p) := -\mathbb{E}\left[\sum_{t=1}^{T} \log p(\mathrm{obs}_t \mid \mathrm{history}_{t-1})\right]$$

Predictive loss controls posterior quality

Sequence prediction loss

a.k.a. neg. log likelihood

$$\ell(p) := -\mathbb{E}\left[\sum_{t=1}^{T} \log p(\mathrm{obs}_t \mid \mathrm{history}_{t-1})\right]$$

Fact (C**N**RZ'25): for any function f,

$$\mathbb{E}\Big[\mathrm{KL}\Big(\mathbb{P}(f(\hat{\tau}_t) \in \cdot\,),\ \mathbb{P}(f(\tau_t^\star) \in \cdot\,) \mid \mathrm{hist}_{t-1}\Big)\Big]$$

$$\leq\ \ell(\hat{p}) - \ell(p^\star) \qquad \text{for any } t = 1, \cdots, T$$

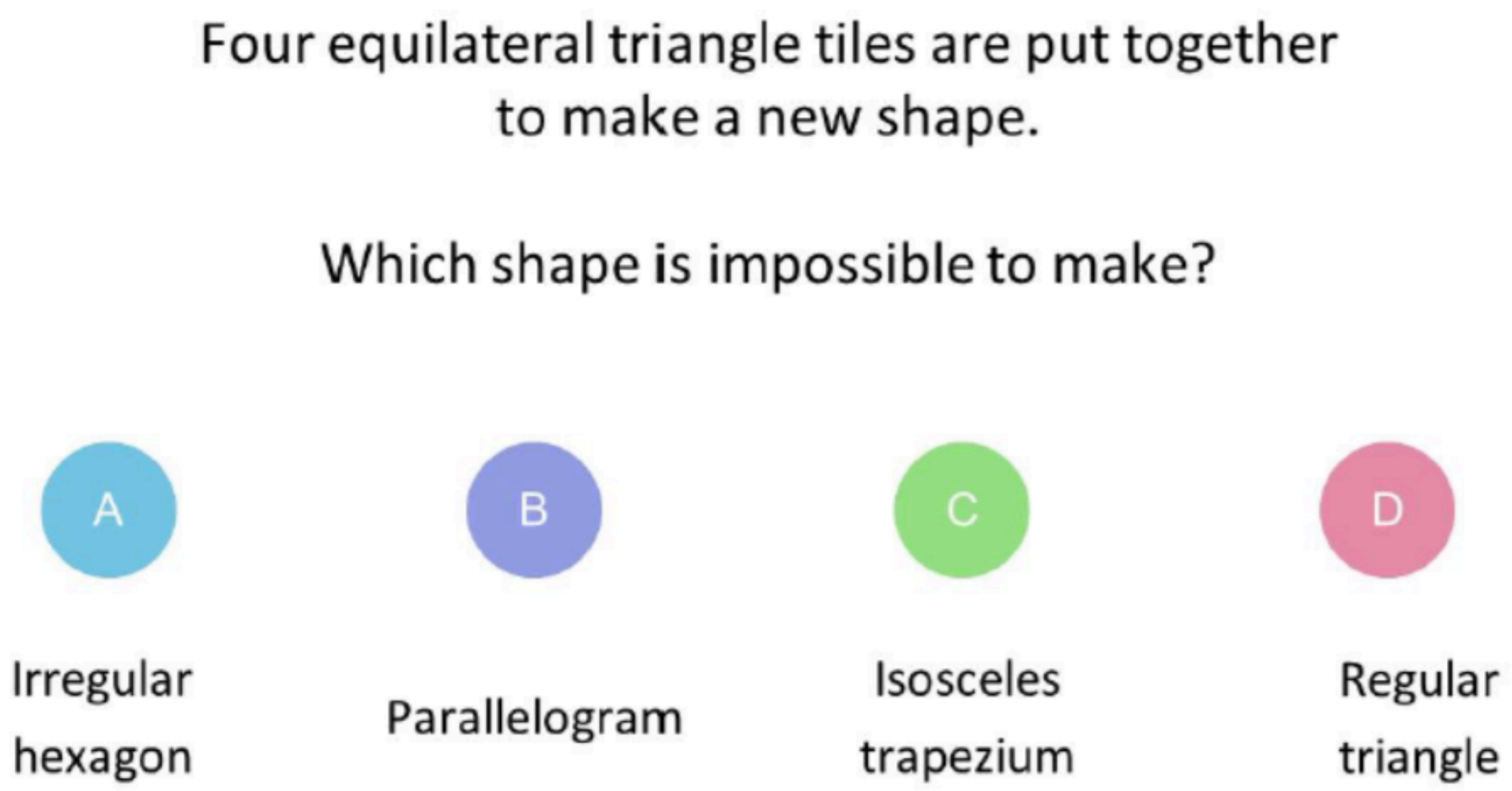
Experiments

Empirical demonstration

Real-world dataset from an online education platform

- Tutoring platform serving millions
- ~1K multiple choice questions on algebra, number theory, and geometry
- Individual responses from students

Example question from EEDI



Diagnostic questions: The NeurIPS 2020 education challenge. https://diagnosticquestions.com/ Wang et al. (2020)

Training a sequence model

- Take LLM pre-trained on internet data;
  already understands language and has world knowledge

Training a sequence model

- Take LLM pre-trained on internet data;
  already understands language and has world knowledge

- Finetune on domain-specific data (usually proprietary)

$$\sum_{\text{students}} \sum_{\text{questions}} \log \widehat{p} \left(\text{answer} \mid \text{question, past Q\&A, student info} \right)$$

- When this data is limited, use parameter-efficient methods, e.g., LoRA

Training a sequence model

- Take LLM pre-trained on internet data;
  already understands language and has world knowledge
- Finetune on domain-specific data (usually proprietary)

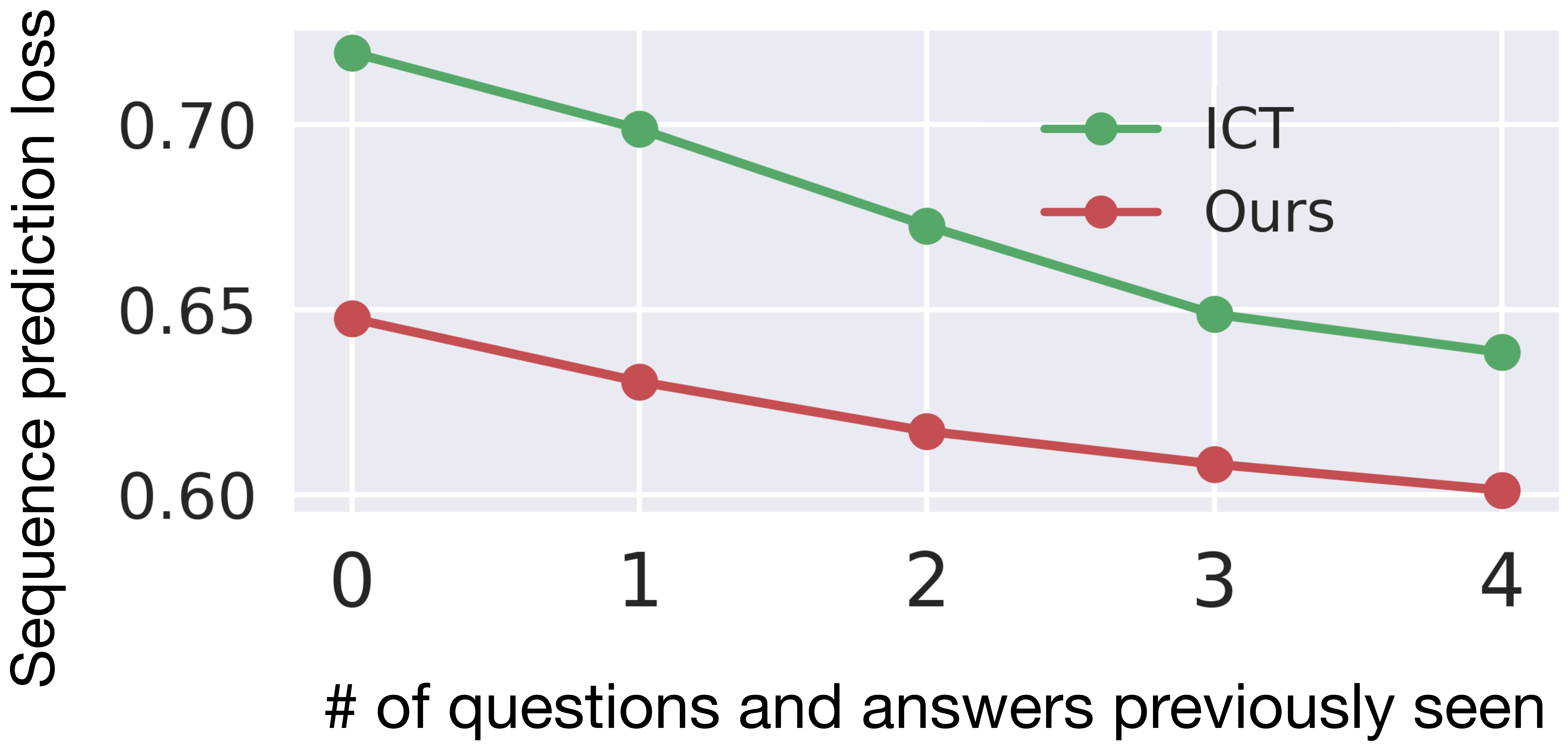
$$\sum_{\text{students}} \sum_{\text{questions}} \log \widehat{p}\left(\text{answer} \mid \text{question, past Q\&A, student info}\right)$$

- When this data is limited, use parameter-efficient methods, e.g., LoRA
- Baseline: Base LLM and in-context training (ICT)

Offline sequence prediction

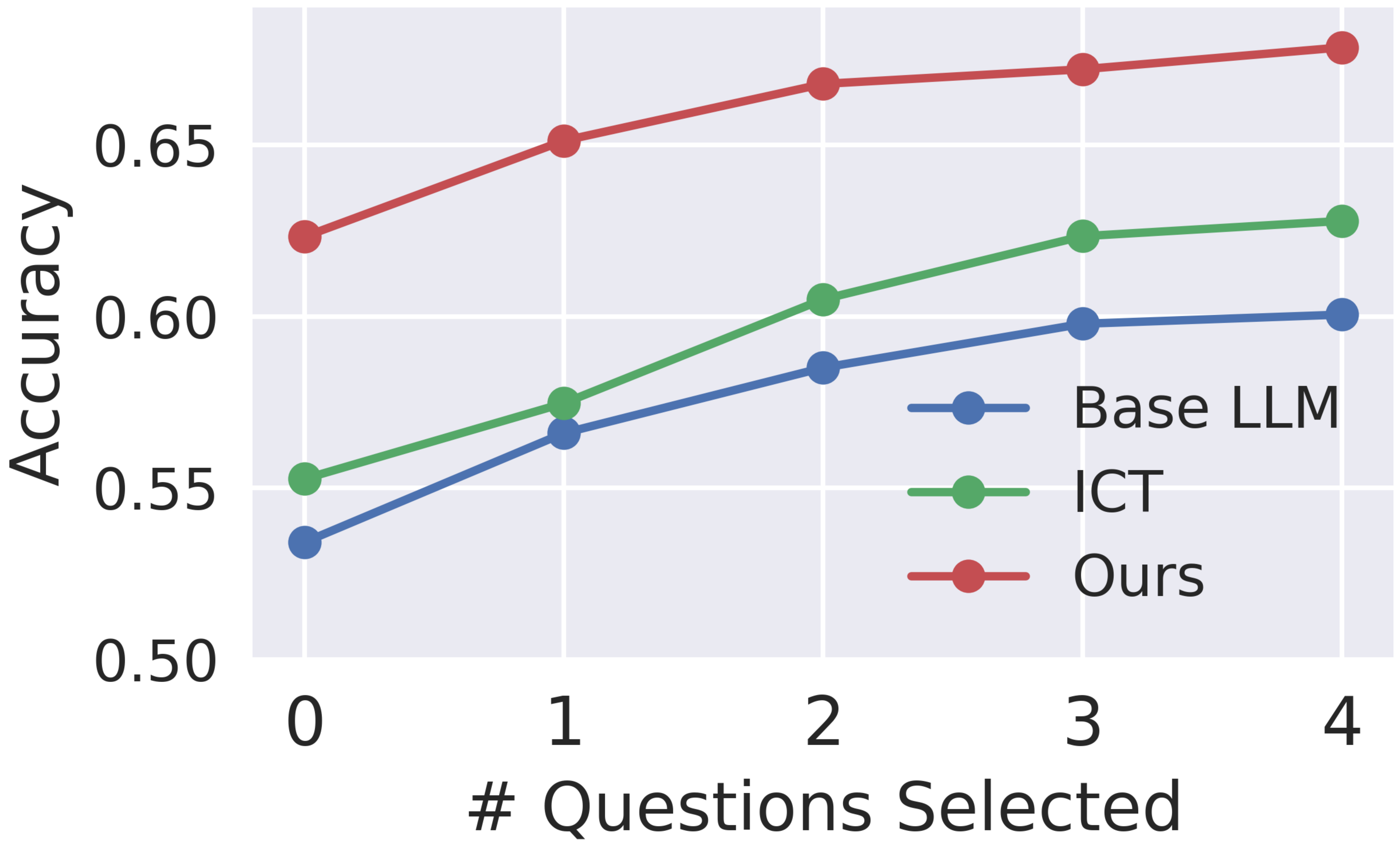
Learning across students

- Llama 3.1 finetuned on EEDI data via LoRA



Offline sequence prediction

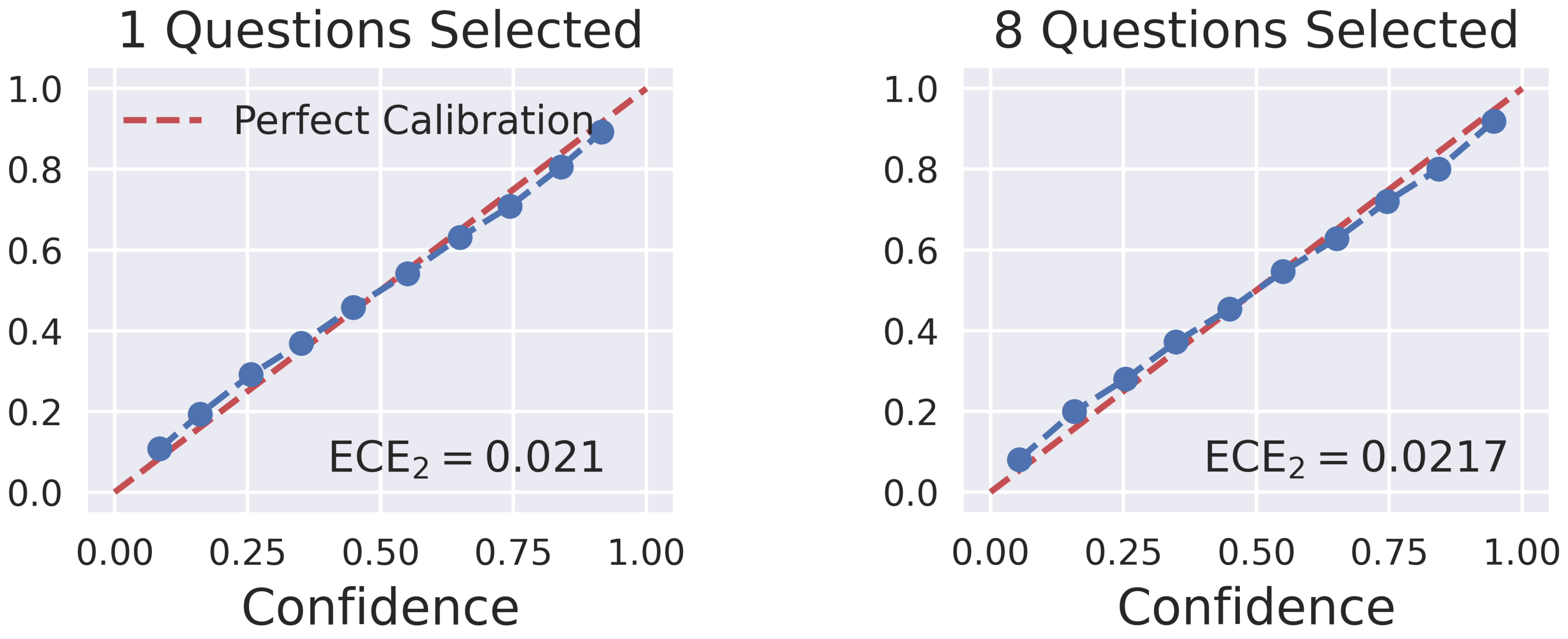
Learning across students



Calibration

Predicted probabilities on student answer sensible?

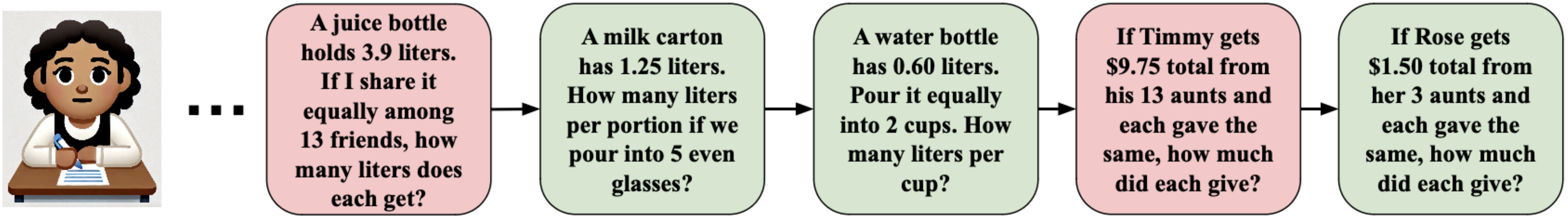
y = Fraction(correct | confidence = x); we want this to be equal to x



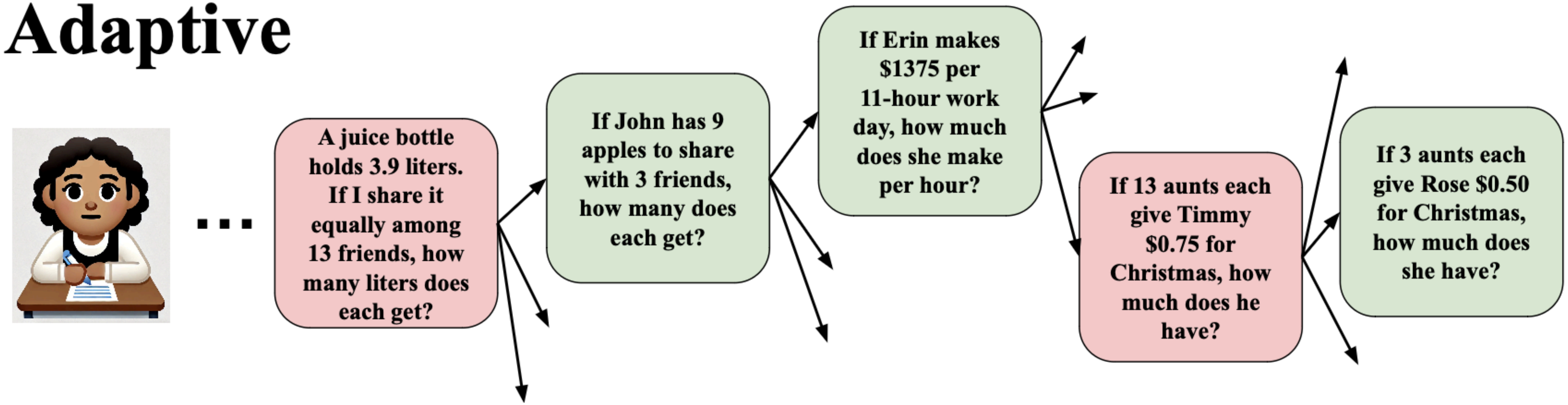
Posteriors in action

Efficient assessment via adaptivity

Static



Adaptive



Markov decision process

Adaptive assessment as sequential decision-making

Predictive view

Reduce uncertainty on future responses as measured by

Entropy(student answers to all questions in problem bank)

Markov decision process

Adaptive assessment as sequential decision-making

Predictive view	Reduce uncertainty on future responses as measured by Entropy(student answers to all questions in problem bank)

Action: which question to pick next?

State: current posterior

Markov decision process

Adaptive assessment as sequential decision-making

Predictive view

Reduce uncertainty on future responses as measured by

Entropy(student answers to all questions in problem bank)

Select Q that can maximally resolve future uncertainty

Action: which question to pick next?

State: current posterior

Markov decision process

Adaptive assessment as sequential decision-making

Predictive view

Reduce uncertainty on future responses as measured by
Entropy(student answers to all questions in problem bank)

Select Q that can maximally resolve future uncertainty

Action: which question to pick next?

State: current posterior

Answer

Markov decision process

Adaptive assessment as sequential decision-making

Predictive view

Reduce uncertainty on future responses as measured by

Entropy(student answers to all questions in problem bank)

Select Q that can maximally resolve future uncertainty

Action: which question to pick next?

State: current posterior

Answer

Reward:
reduction in entropy

Markov decision process

Adaptive assessment as sequential decision-making

Predictive view

Reduce uncertainty on future responses as measured by
Entropy(student answers to all questions in problem bank)

Select Q that can maximally resolve future uncertainty

Action: which question to pick next?

State: current posterior

Answer

Reward:
reduction in entropy

Answer appended to LLM prompt (“state transition”)

Markov decision process

Adaptive assessment as sequential decision-making

Predictive view	Reduce uncertainty on future responses as measured by Entropy(student answers to all questions in problem bank)

To (approximately) solve this MDP, the AI must implicitly

1. form informed prior based on information on the student
2. use question and solution to comprehend areas of largest uncertainty
3. select questions balancing exploration & exploitation
4. sharpen beliefs based on the student's answer

Markov decision process

Adaptive assessment as sequential decision-making

Rewards (entropy reduction) computable only through our sequence model

Markov decision process

Adaptive assessment as sequential decision-making

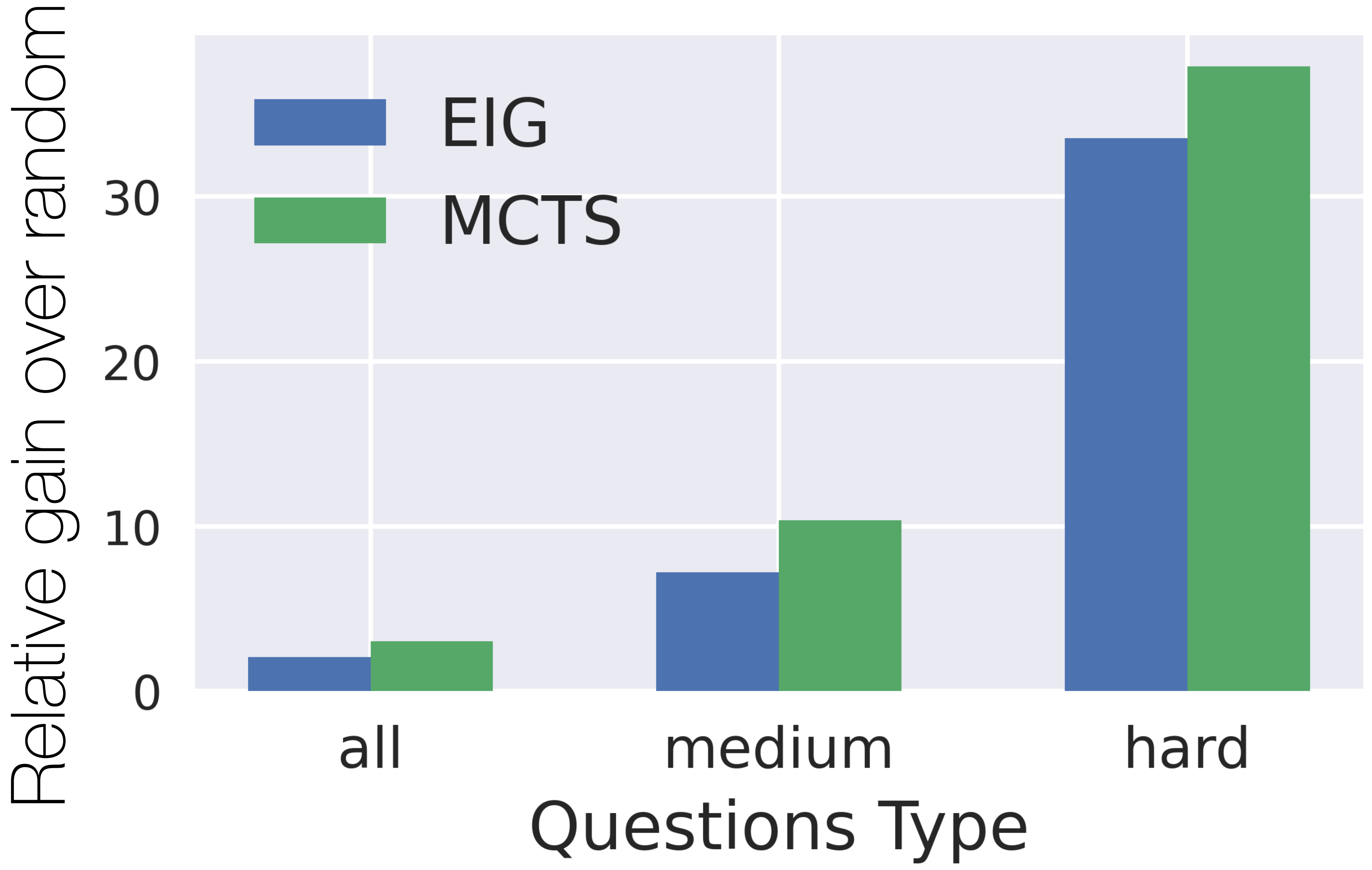
Rewards (entropy reduction) computable only through our sequence model

- MDP formulation allows applying any ADP principle
- Today: basic strategies
  - Expected Information Gain; a.k.a. greedy
  - Monte Carlo Tree Search; sophisticated planning

    Chang, Fu, Hu, Marcus (2005). "An Adaptive Sampling Algorithm for Solving Markov Decision Processes"

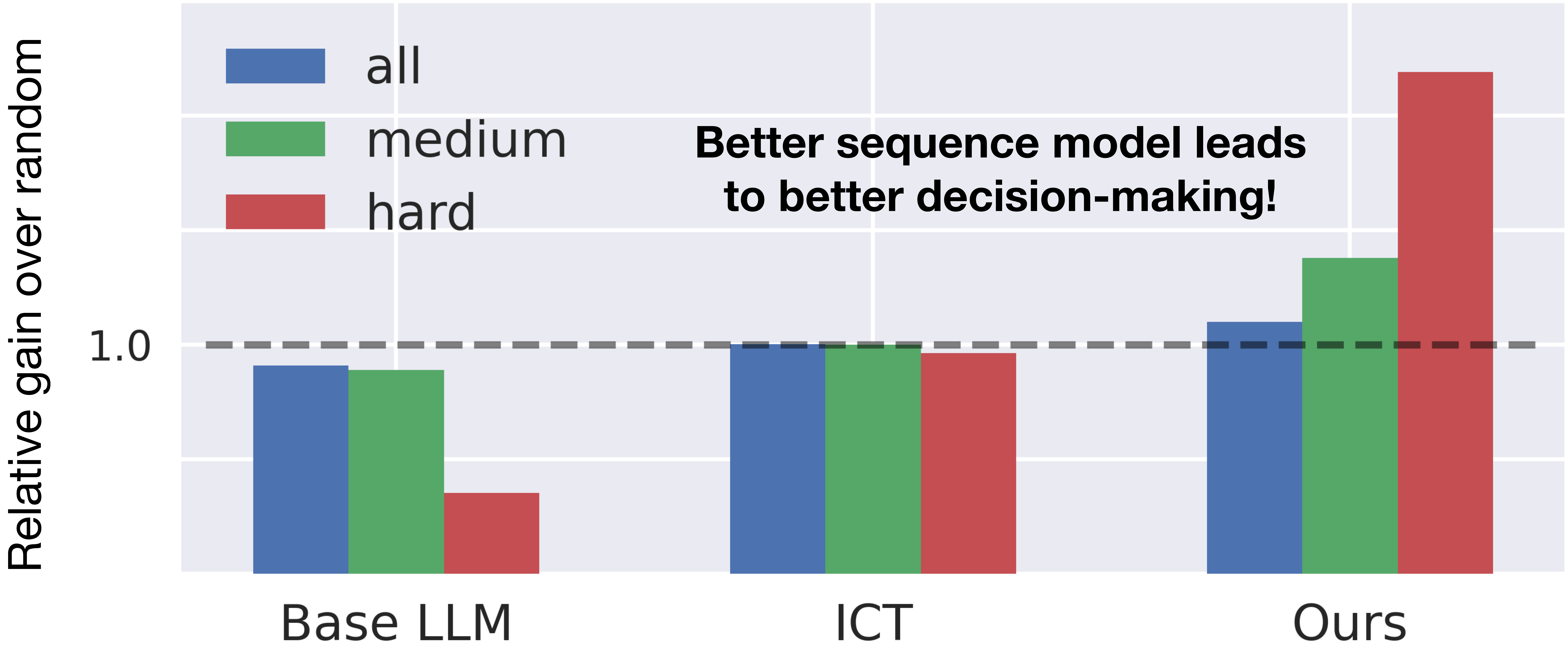
When is Adaptivity Most Helpful?

- Hard to learn that a student is struggling in an area where other students generally do not
- Evaluate uncertainty reduction on subgroups of problem bank
    - student's answer chosen by less than either 50% ("medium") or 30% ("hard") across the population.
- Adaptivity helps by selecting a test question that most find easy but this student may answer incorrectly



Sequence prediction & interactive decisions

Comparison of sequence models

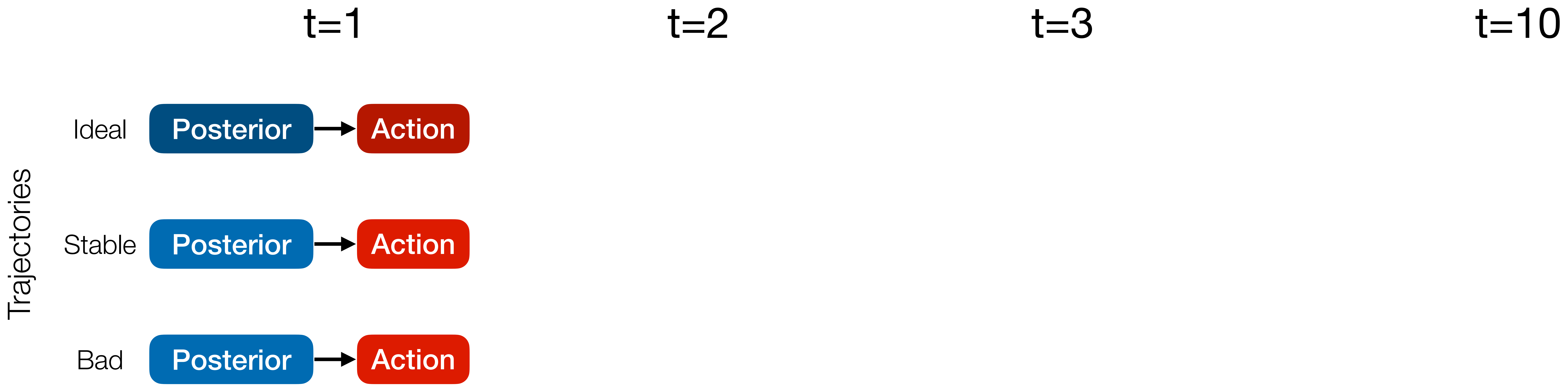


Theoretical insights

Progress in LLM pre-training directly improves interactive decision-making performance

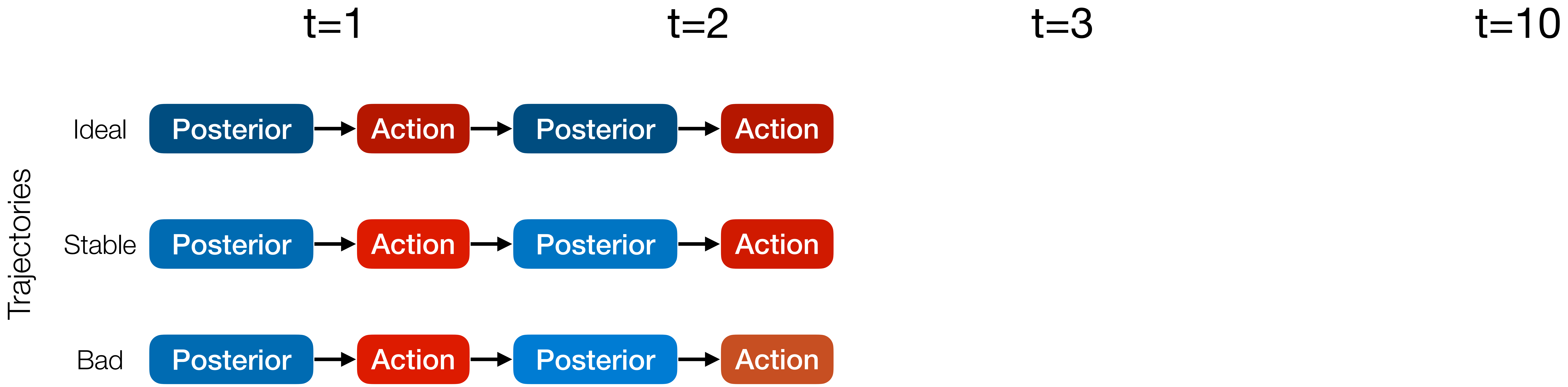
Offline predictions to online decisions

Do initial mistakes magnify over time?



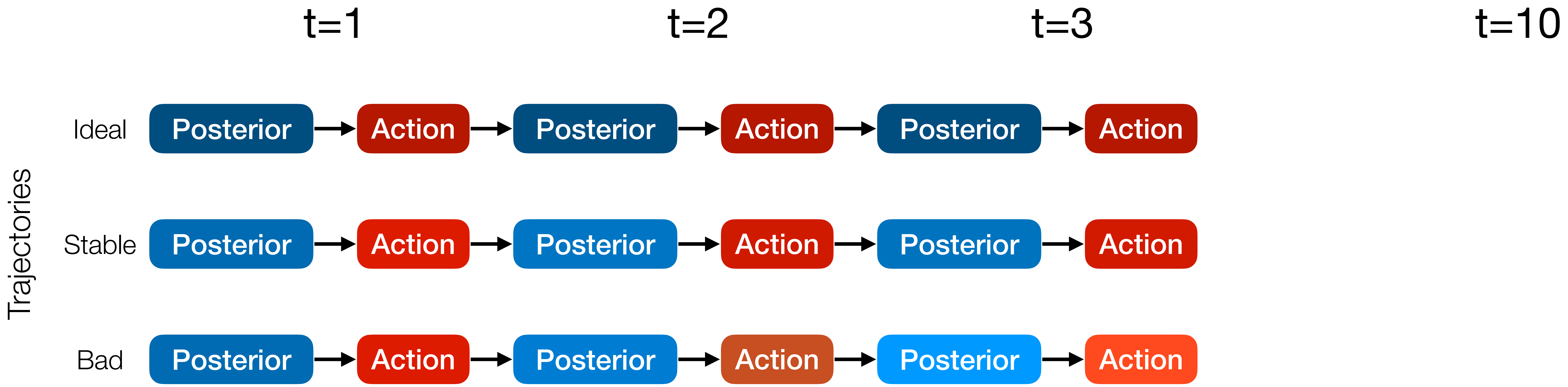
Offline predictions to online decisions

Do initial mistakes magnify over time?



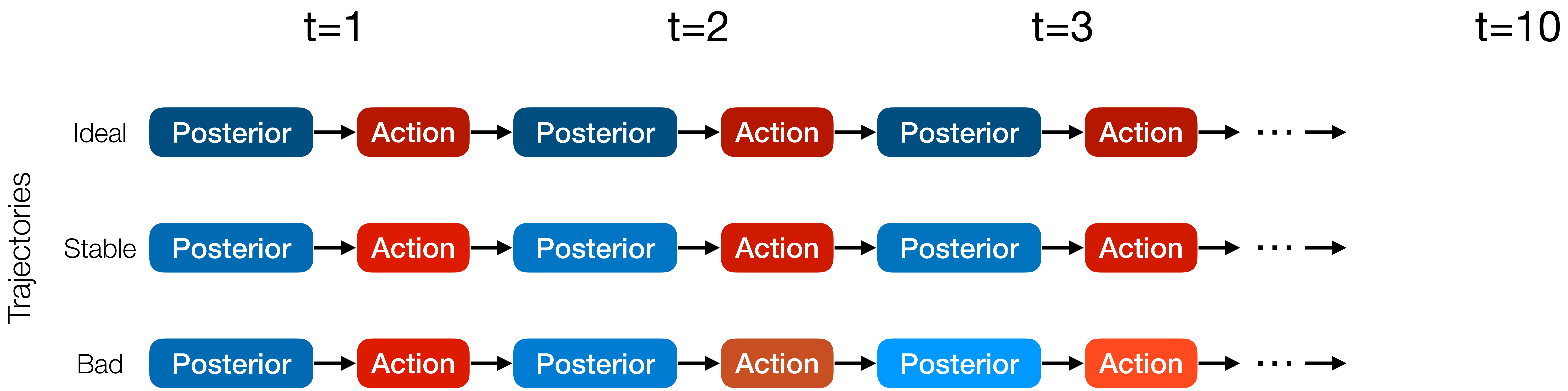
Offline predictions to online decisions

Do initial mistakes magnify over time?



Offline predictions to online decisions

Do initial mistakes magnify over time?



Offline predictions to online decisions

Do initial mistakes magnify over time?

Offline predictions to online decisions

Do initial mistakes magnify over time?

Will small initial prediction errors lead to cascading decisions over time?
Simchowitz et al. ('20) showed yes in the **worst-case**

Offline predictions to online decisions

Do initial mistakes magnify over time?

Main theoretical contribution

- On average, initially small disparities do **NOT** magnify over time
- Dispels pessimistic results on misspecification in the literature

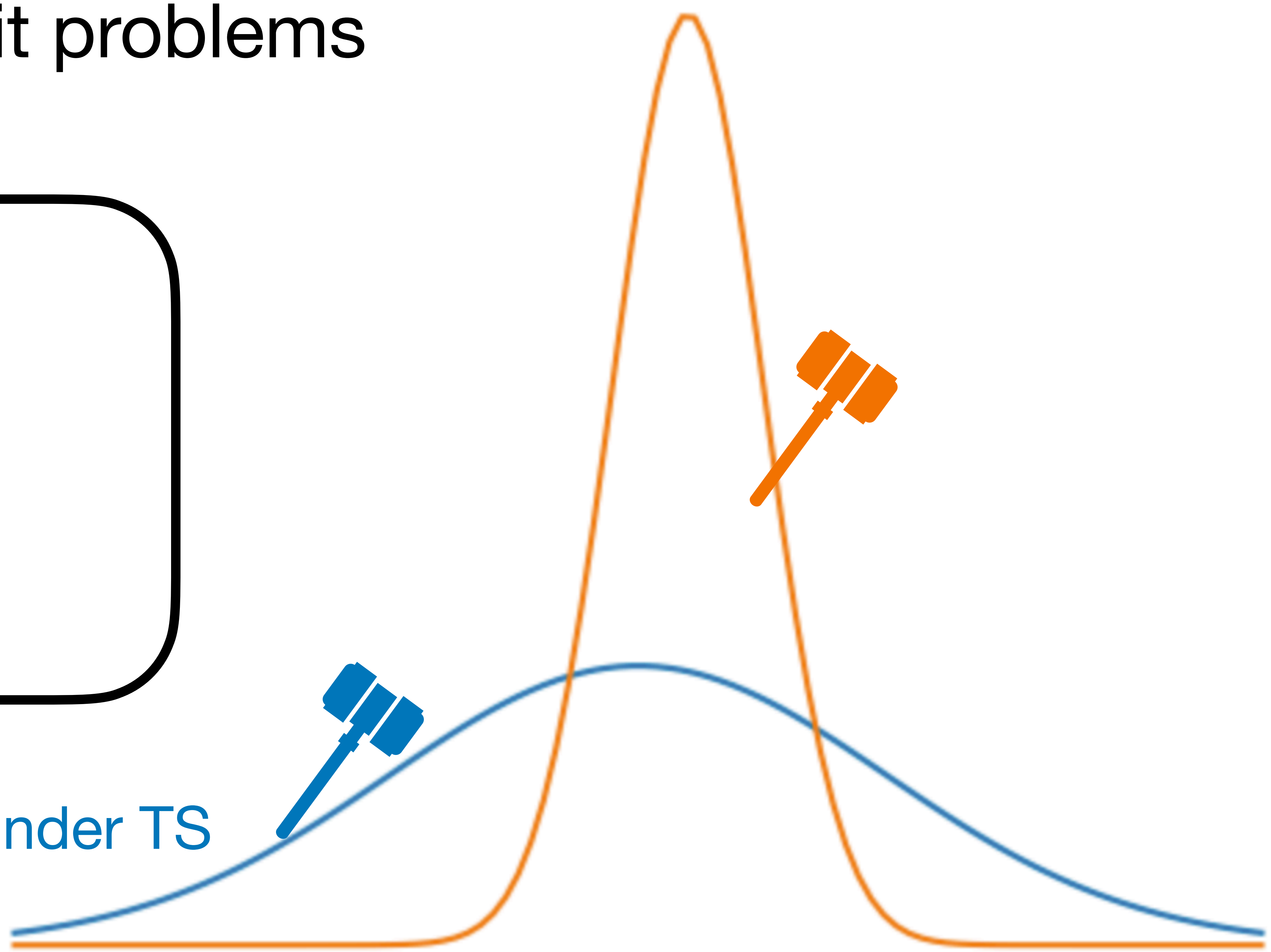
Bandits with language-based action features

Simplified setting for illustration

- MDPs hard to analyze; history generated by our policy not oracle
- Study how our framework performs in bandit problems

Policy: Thomson sampling

Draw fictitious future from posterior,
pick the best arm that gives best future



Blue arm has chance of being explored under TS

Vignette: Thomson sampling

Autoregressive generation reveals actions that *might* have great performance

1) Fill in missing outcomes by autoregressive generation

Observed and 'missing' outcomes

Potential rewards

$Z^1 \quad Y_1^1, Y_2^1, \hat{Y}_3^1, \cdots, \hat{Y}_T^1$

$\vdots$

$Z^k \quad Y_1^k, \hat{Y}_2^k, \hat{Y}_3^k, \cdots, \hat{Y}_T^k$

Vignette: Thomson sampling

Autoregressive generation reveals actions that *might* have great performance

1) Fill in missing outcomes by autoregressive generation

Observed and 'missing' outcomes

Potential rewards

$Z^1 \quad Y_1^1, Y_2^1, \hat{Y}_3^1, \cdots, \hat{Y}_T^1$

$\vdots$

$Z^k \quad Y_1^k, \hat{Y}_2^k, \hat{Y}_3^k, \cdots, \hat{Y}_T^k$

2) Compute reward rates under hypothetical table

Average

$$\hat{\mu}^1 = \frac{Y_1^1 + Y_2^1 + \hat{Y}_3^1 + \cdots + \hat{Y}_T^1}{T}$$

Average

$$\hat{\mu}^k = \frac{Y_1^k + \hat{Y}_2^k + \hat{Y}_3^k + \cdots + \hat{Y}_T^k}{T}$$

Vignette: Thomson sampling

Autoregressive generation reveals actions that *might* have great performance

1) Fill in missing outcomes by autoregressive generation

2) Compute reward rates under hypothetical table

3) Pick arm with highest hypothetical reward

Observed and 'missing' outcomes

Potential rewards

$Z^1 \quad Y_1^1, Y_2^1, \hat{Y}_3^1, \cdots, \hat{Y}_T^1$

$\vdots$

$Z^k \quad Y_1^k, \hat{Y}_2^k, \hat{Y}_3^k, \cdots, \hat{Y}_T^k$

Average

$$\hat{\mu}^1 = \frac{Y_1^1 + Y_2^1 + \hat{Y}_3^1 + \cdots + \hat{Y}_T^1}{T}$$

Average

$$\hat{\mu}^k = \frac{Y_1^k + \hat{Y}_2^k + \hat{Y}_3^k + \cdots + \hat{Y}_T^k}{T}$$

argmax

Data assumptions

1. **Independently drawn articles.**
   - Text/ potential outcome sets $\left(Z^{(a)}, Y_1^{(a)}, \dots, Y_T^{(a)}\right)$ are independent across articles.

2. Historical data is representative of future days

- Distribution $\left(Z^{(a)}, Y_1^{(a)}, \dots, Y_T^{(a)}\right)$ is the same for articles in historical data as what governs tomorrow's draw..

3. Exchangeability across users

- $P^*\left(Y_1^{(a)}, \dots, Y_T^{(a)} \mid Z^{(a)}\right) = P^*\left(Y_{\sigma(1)}^{(a)}, \dots, Y_{\sigma(T)}^{(a)} \mid Z^{(a)}\right)$

Regret controlled by sequence loss

Sequence prediction loss $\quad \ell_T(p) := -\mathbb{E}\Big[\sum_{\text{T questions}} \log p\,(\text{answer} \mid \text{question, past data})\Big]$

Theorem (C**N**RZ'24)

Impossible to implement

Val loss

$$\begin{array}{c}\text{Regret of TS}\\ \text{with model } \widehat{p}\end{array} \leq \begin{array}{c}\text{Regret of oracle TS}\\ \text{with true model } p^\star\end{array} + \sqrt{2 \cdot \text{no. actions} \cdot (\ell_T(\widehat{p}) - \ell_T(p^\star))}$$

Regret controlled by sequence loss

Sequence prediction loss $\ell_T(p) := -\mathbb{E}\Big[\sum_{\text{T questions}} \log p\,(\text{answer} \mid \text{question, past data})\Big]$

Theorem (C**N**RZ'24)

Impossible to implement

Val loss

Regret of TS with model $\widehat{p}$ $\leq$ Regret of oracle TS with true model $p^\star$ $+ \sqrt{2 \cdot \text{no. actions} \cdot (\ell_T(\widehat{p}) - \ell_T(p^\star))}$

Scaling laws govern online decision-making performance!

Regret controlled by sequence loss

Sequence prediction loss $\quad \ell_T(p) := -\mathbb{E}\Big[\sum_{\text{T questions}} \log p\,(\text{answer} \mid \text{question, past data})\Big]$

Theorem (C**N**RZ'24)

Impossible to implement

Val loss

$$\text{Regret of TS with model } \hat{p} \leq \text{Regret of oracle TS with true model } p^\star + \sqrt{2 \cdot \text{no. actions} \cdot (\ell_T(\hat{p}) - \ell_T(p^\star))}$$

Theorem (Y**N**'24) $\quad \ell_T(\hat{p}) - \ell_T(p^\star) \sim \log T$ for reasonable models

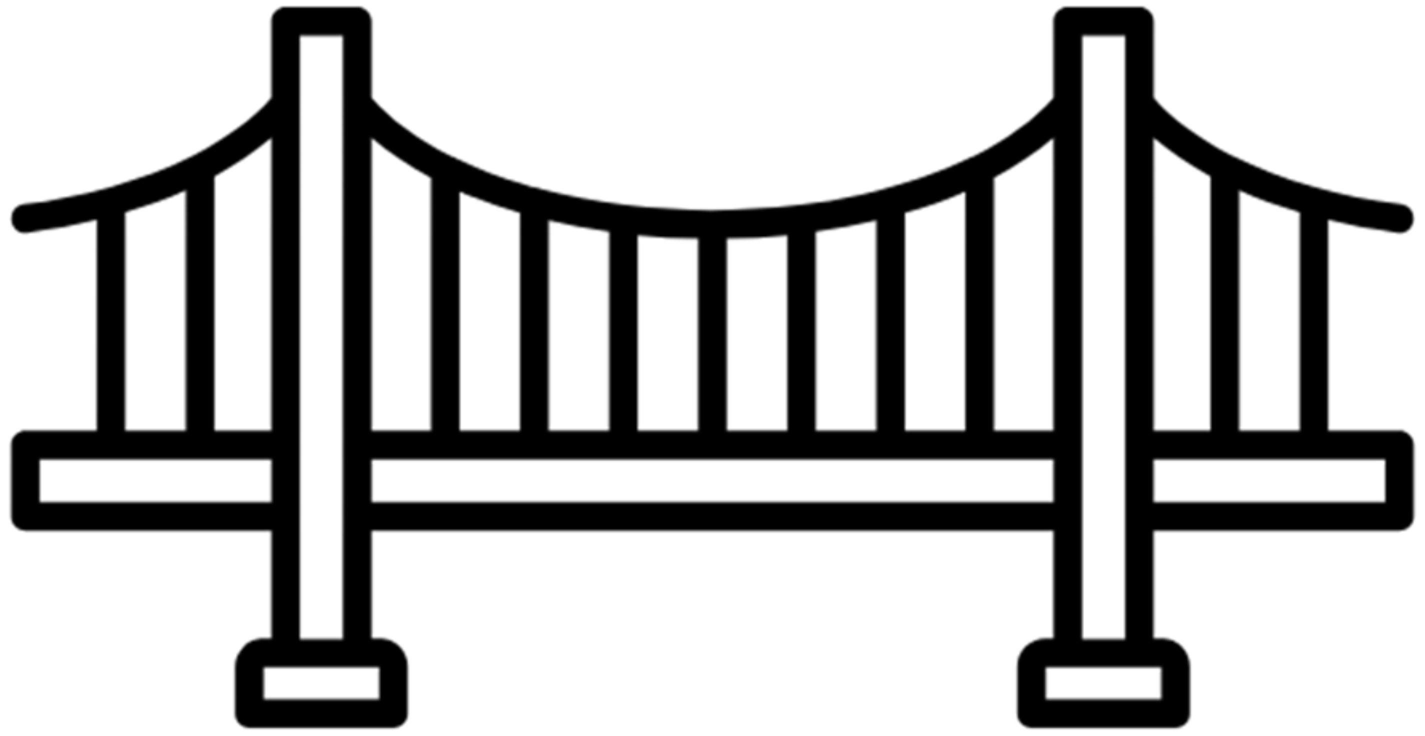
Summary of contributions

- Algorithm: autoreg. generation quantifies uncertainty from missing data
- Theory: accurate offline sequence modeling implies low regret
- Experiments: scalable implementations with LLMs.

Summary of contributions

- Algorithm: autoreg. generation quantifies uncertainty from missing data
- Theory: accurate offline sequence modeling implies low regret
- Experiments: scalable implementations with LLMs.

Sequence Modeling



Uncertainty Quantification

References

- [Uncertainty as missing data] Active Exploration via Autoregressive Generation of Missing Data, Cai, Namkoong, Russo, and Zhang (2024)
- [Connection to Bayesian modeling] Exchangeable Sequence Models Quantify Uncertainty Over Latent Concepts, Ye and Namkoong (2024)
- [Language experiments] Adaptive Elicitation of Latent Information Using Natural Language, Wang, Zollo, Zemel, and Namkoong (2025)
- [Contextual bandits] Contextual Thompson Sampling via Generation of Missing Data, Zhang, Cai, Namkoong, and Russo (2024)
- [NN Architectures] Architectural and Inferential Inductive Biases For Exchangeable Sequence Modeling, Mittal, Li, Yen, Guetta, and Namkoong (2025)

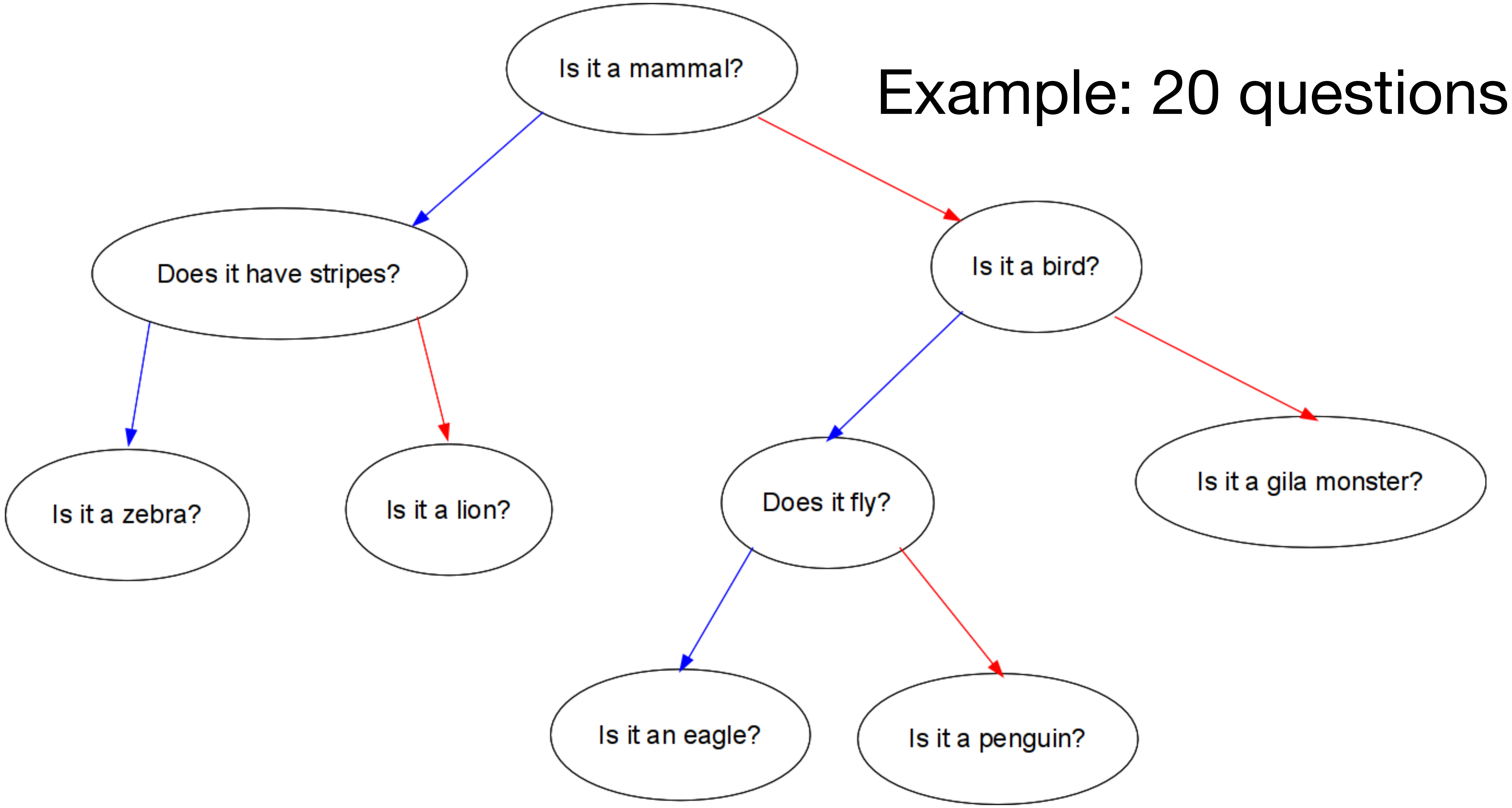
What's next?

Replace human intelligence in algorithm design

Step 1: Data curation

- Come up with tasks that require articulating uncertainty and acting to resolve it
- Key requirement: any task-generation must be applicable to web-scale data

Example: 20 questions



What's next?

Replace human intelligence in algorithm design

Step 1: Data curation

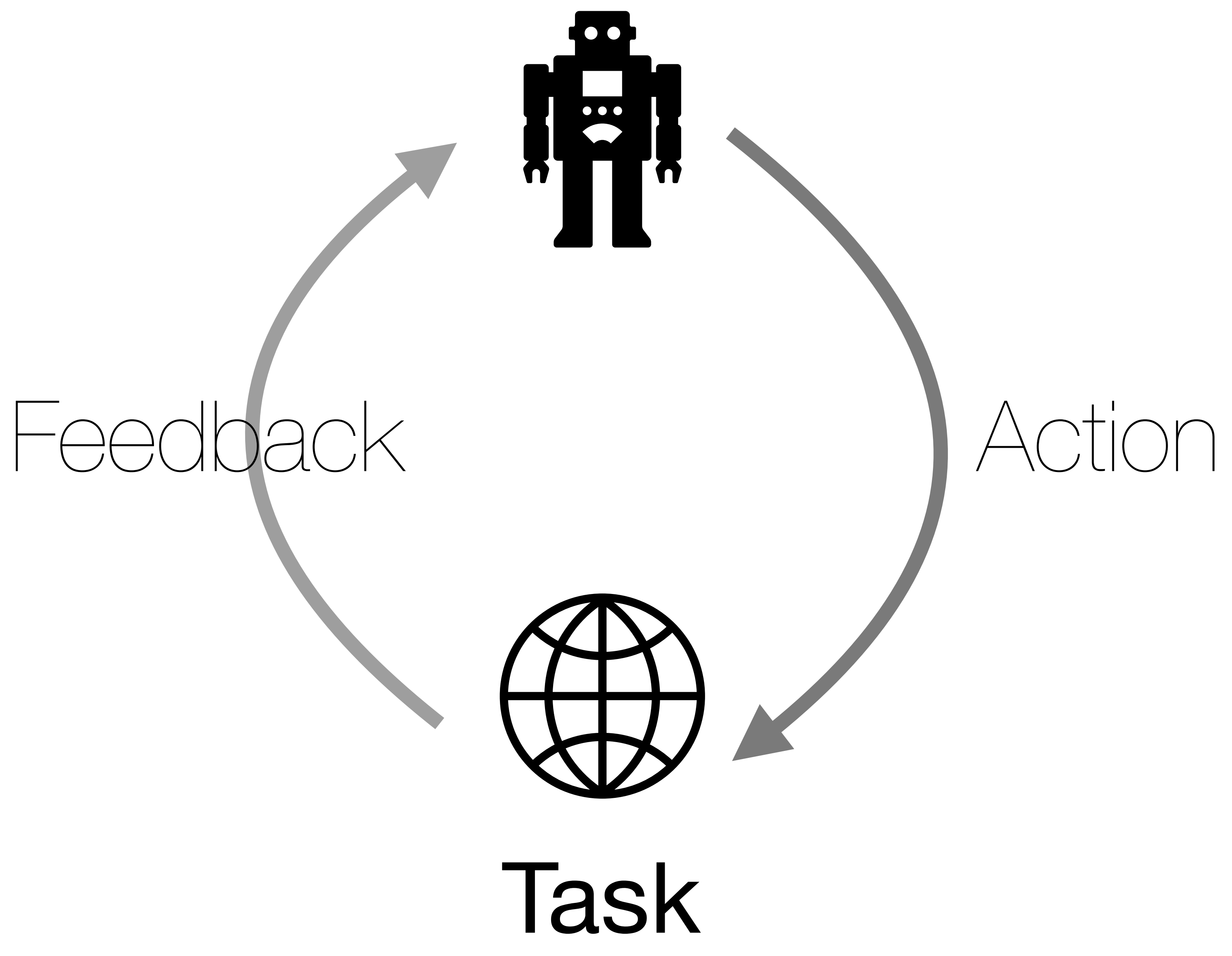
- Come up with tasks that require articulating uncertainty and acting to resolve it
- Key requirement: any task-generation must be applicable to web-scale data

Step 2: Training

- Don't rely on human intelligence to come up with good "learning algorithms"
- Instead, let AI learn to learn through experience

"RL Algorithm"

LLM(action | history, task)



What's next?

Replace human intelligence in algorithm design

Step 1: Data curation

- Come up with tasks that require articulating uncertainty and acting to resolve it
- Key requirement: any task-generation must be applicable to web-scale data

Step 2: Training

- Don't rely on human intelligence to come up with good "learning algorithms"
- Instead, let AI learn to learn through experience

Step 3: Scaling

- Our job is to figure out a way to generate billions of experiences
- Study scaling behavior on internet-scale data