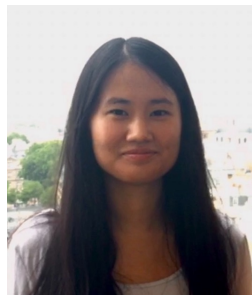
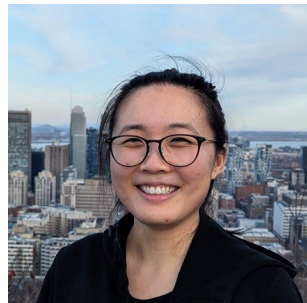


Posterior Sampling via Autoregressive Generation

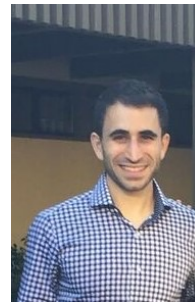
Hong Namkoong
Columbia University



Kelly W. Zhang



Tiffany Cai



Daniel Russo







Goal: AI-driven decisions

- AI now comprehends language and visual inputs
- Big opportunities to make **decisions** based on them
- Decision-making requires comprehending **uncertainty** and acting to resolve it

Job recommendations



Top job picks for you
Based on your profile and search history

-  **Distinguished Applied Researcher** ✓ ×
Capital One · San Francisco, CA
 166 school alumni work here
Promoted
-  **Applied Researcher II** ✓ ×
Capital One · San Francisco, CA
 61 company alumni work here
Promoted · **Be an early applicant**
-  **Member of Research Staff - Fujitsu Research** ✓ ×
Fujitsu · Sunnyvale, CA (On-site)
 1 company alum works here
Promoted · **Be an early applicant**

[Show all →](#)

Cold-start a notorious problem in RecSys

Longstanding challenge in AI

Despite *many* attempts, neural nets still cannot comprehend uncertainty

We know two things in AI



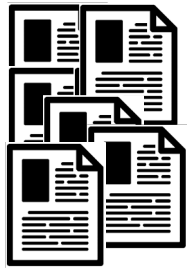
1. Scalable optimization (a.k.a. auto-differentiation)
2. Rigorous empirical validation based on OOS loss

Wait what about {Bayesian NNs, deep GPs, conformal prediction, ensembling etc}??

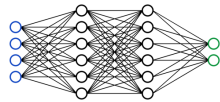
TLDR; very challenging to employ above

Online decision-making within a day

New articles are released



An LLM reads them



Interact with User 1

1) Action / rec:							
2) Observation:	<table><tr><td>Click</td><td>✓</td></tr><tr><td>Like</td><td>✗</td></tr><tr><td>Share</td><td>✓</td></tr></table>	Click	✓	Like	✗	Share	✓
Click	✓						
Like	✗						
Share	✓						
3) Reward:							



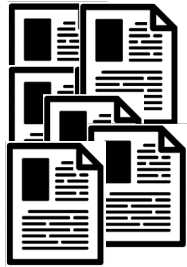
Interact with User T

1) Action / rec:							
2) Observation:	<table><tr><td>Click</td><td>✗</td></tr><tr><td>Like</td><td>✗</td></tr><tr><td>Share</td><td>✗</td></tr></table>	Click	✗	Like	✗	Share	✗
Click	✗						
Like	✗						
Share	✗						
3) Reward:							

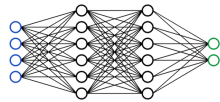



Online decision-making within a day



New articles are released




An LLM reads them





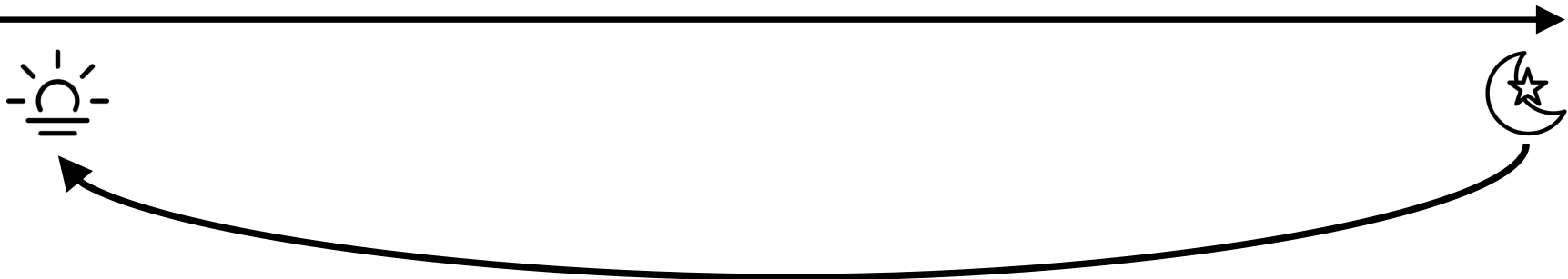
 Interact with User 1

1) Action / rec:							
2) Observation:	<table><tr><td>Click</td><td>✓</td></tr><tr><td>Like</td><td>✗</td></tr><tr><td>Share</td><td>✓</td></tr></table>	Click	✓	Like	✗	Share	✓
Click	✓						
Like	✗						
Share	✓						
3) Reward:							



 Interact with User T

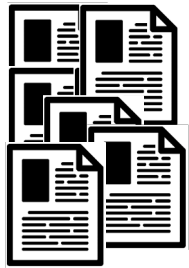
1) Action / rec:							
2) Observation:	<table><tr><td>Click</td><td>✗</td></tr><tr><td>Like</td><td>✗</td></tr><tr><td>Share</td><td>✗</td></tr></table>	Click	✗	Like	✗	Share	✗
Click	✗						
Like	✗						
Share	✗						
3) Reward:							



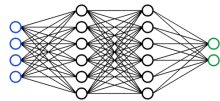
Repeat process tomorrow


Prior art

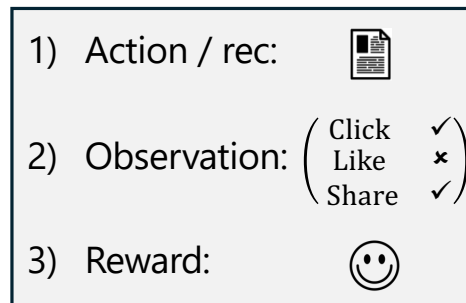
New articles are released




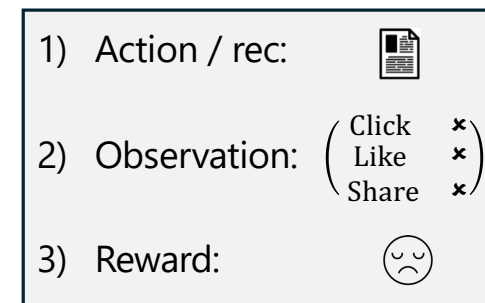
An LLM reads them



 Interact with User 1



 Interact with User T



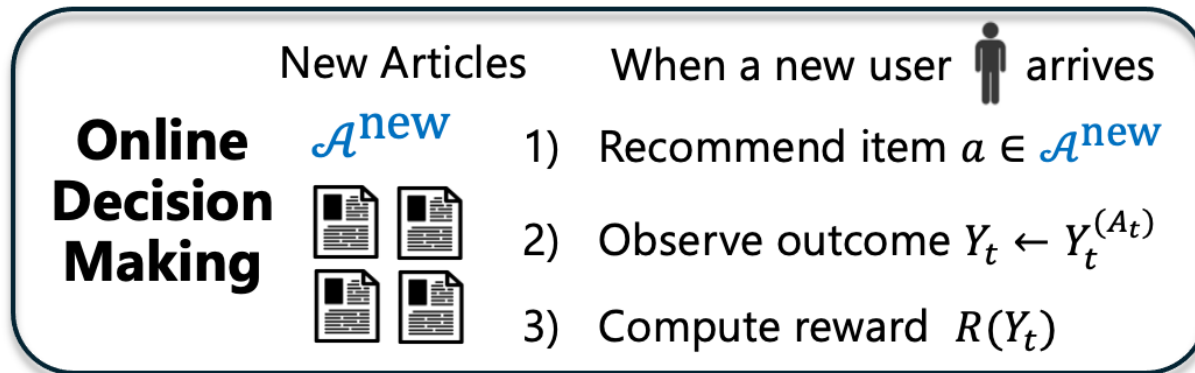
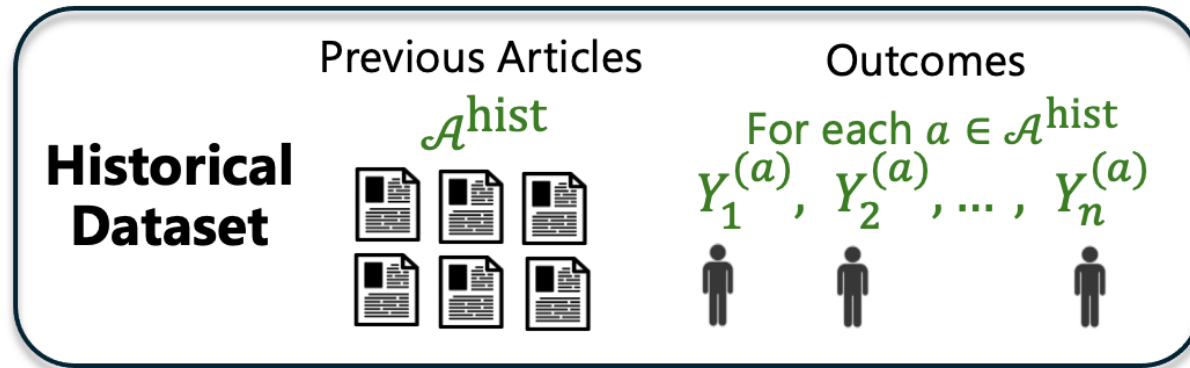
Current SoTA: Thomson sampling with uninformative prior, e.g., based on article categories

Today

1. Extract 'informed prior' from LLM
2. Comprehend remaining uncertainty
3. Balance exploration / exploitation

For this 20min talk, I focus on non-contextual MAB setting for simplicity;
main insight generalizes to contextual problems

Today: learning to learn



Use vast historical data to **warm-start** online decision-making

Thomson sampling



Z : LLM features

U : other latent factors that govern article popularity

- Draw U from the posterior given all data about the article
- Pick best article according to the drawn values

This requires positing a prior and doing posterior updates over LLM features Z —longstanding challenge

Main insight due to De Finetti (1930)



Z : LLM features

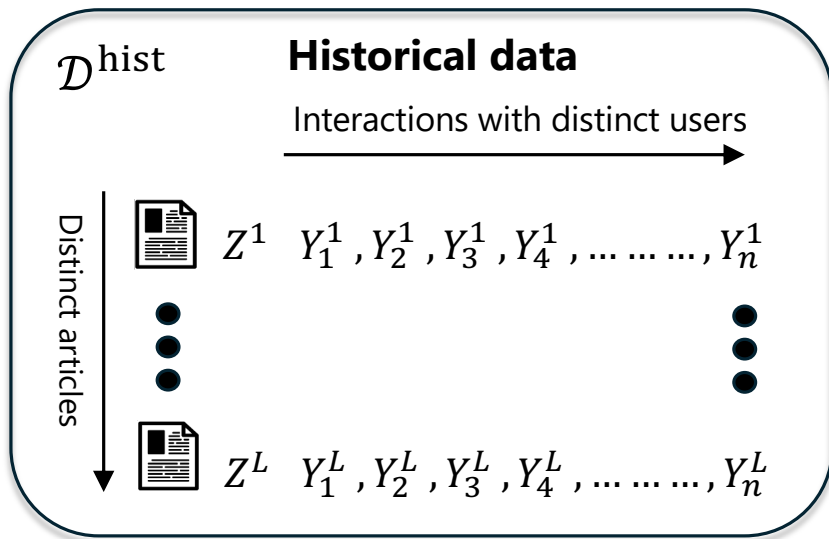
U : other latent factors that govern article popularity

Autoregressive modeling of exchangeable observations implicitly learns a Bayesian model

$$\text{maximize} \quad \sum_{t=1}^n \log p_{\theta}(Y_t^{(a)} \mid Z^{(a)}, Y_{1:t-1}^{(a)})$$

1) Pretrain a sequence model on historical data

To attain low loss, one must *implicitly* comprehend uncertainty given text/interaction data

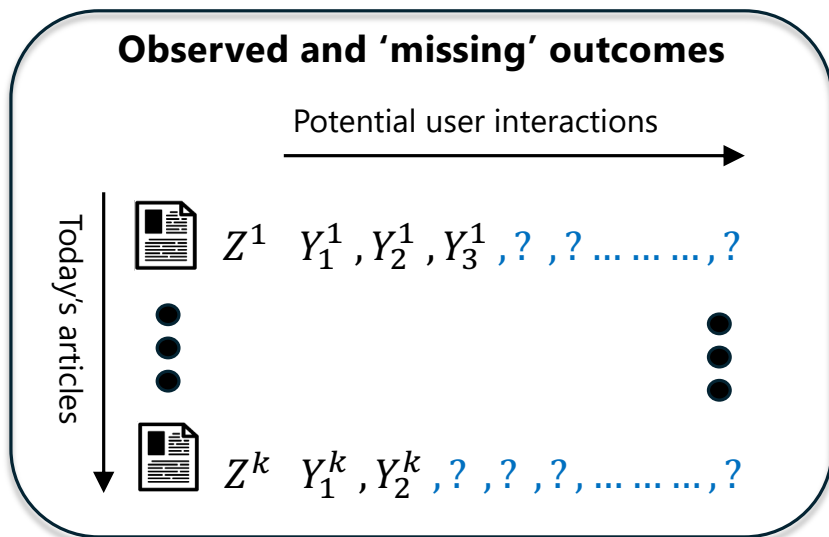


Train transformer on the usual sequence loss

$$\sum_{t=1}^n \log p_{\theta}(Y_t^{(a)} \mid Z^{(a)}, Y_{1:t-1}^{(a)})$$

2) Act on uncertainty by autoregressive generation

Sampling hypothetical outcomes reveals actions that *might* have great performance

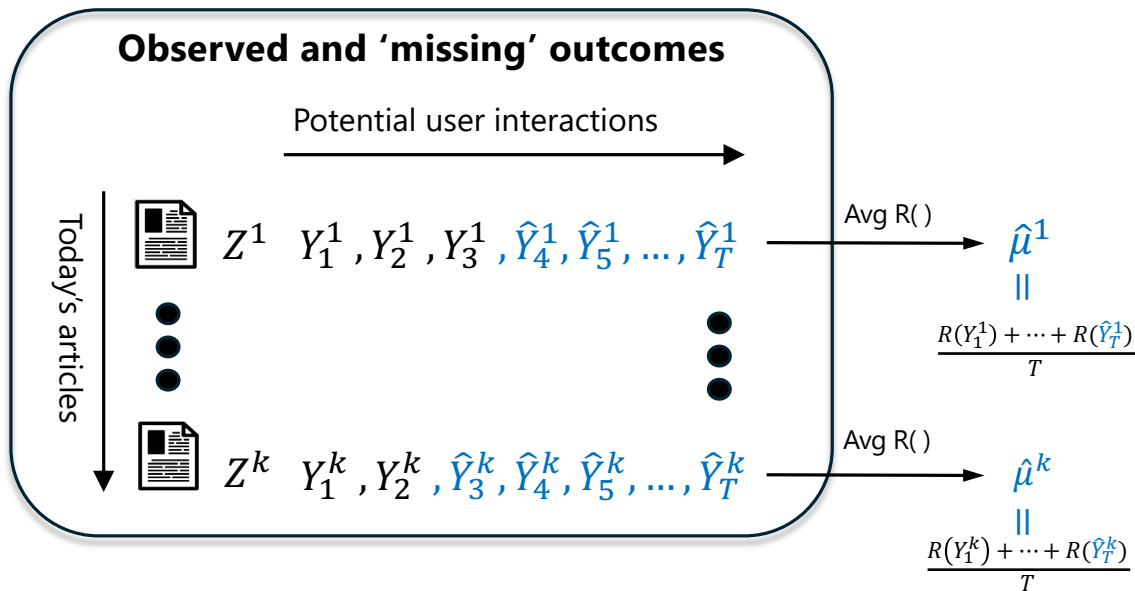


2) Act on uncertainty by autoregressive generation

Sampling hypothetical outcomes reveals actions that *might* have great performance

1) Fill in missing outcomes by autoregressive generation

2) Compute reward rates under hypothetical table



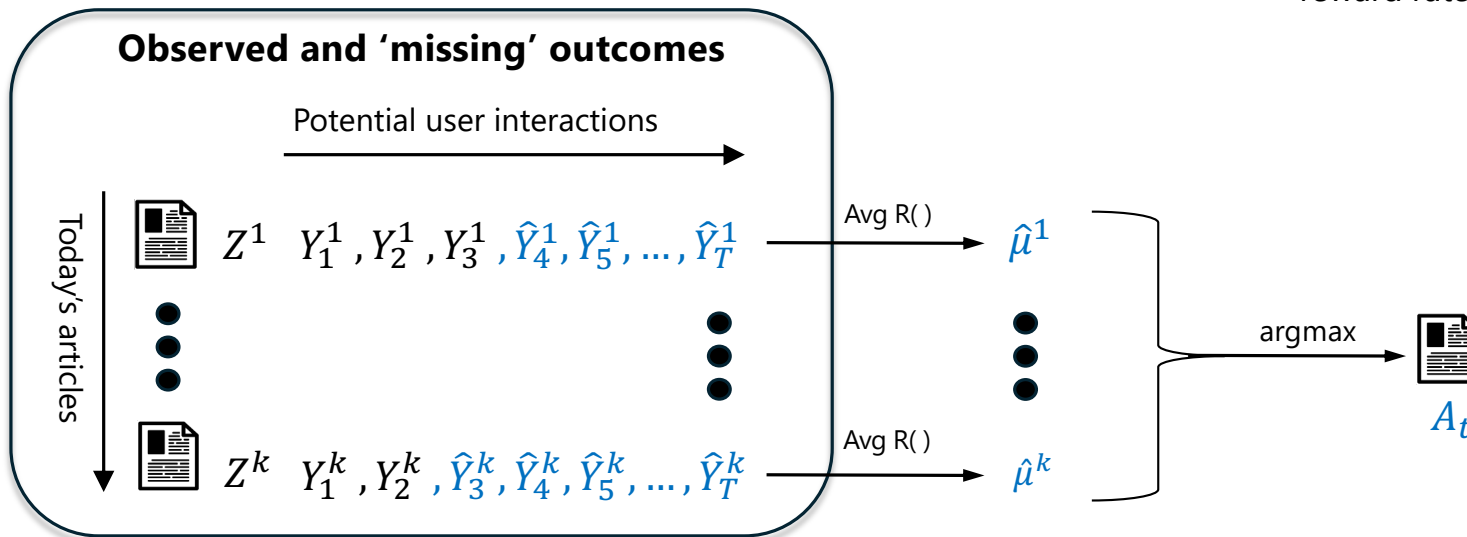
2) Act on uncertainty by autoregressive generation

Sampling hypothetical outcomes reveals actions that *might* have great performance

1) Fill in missing outcomes by autoregressive generation

2) Compute reward rates under hypothetical table

3) Pick the item with highest hypothetical reward rate



2) Act on uncertainty by autoregressive generation

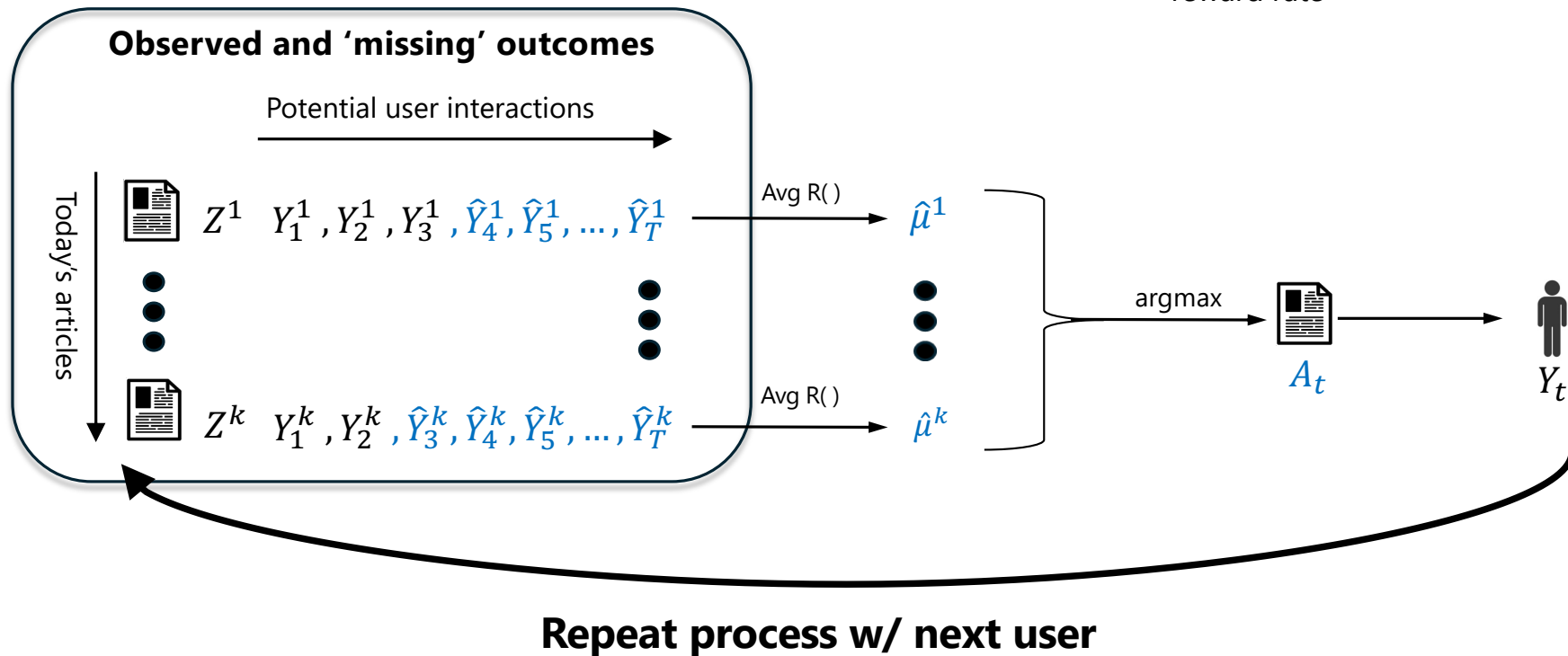
Sampling hypothetical outcomes reveals actions that *might* have great performance

1) Fill in missing outcomes by autoregressive generation

2) Compute reward rates under hypothetical table

3) Pick the item with highest hypothetical reward rate

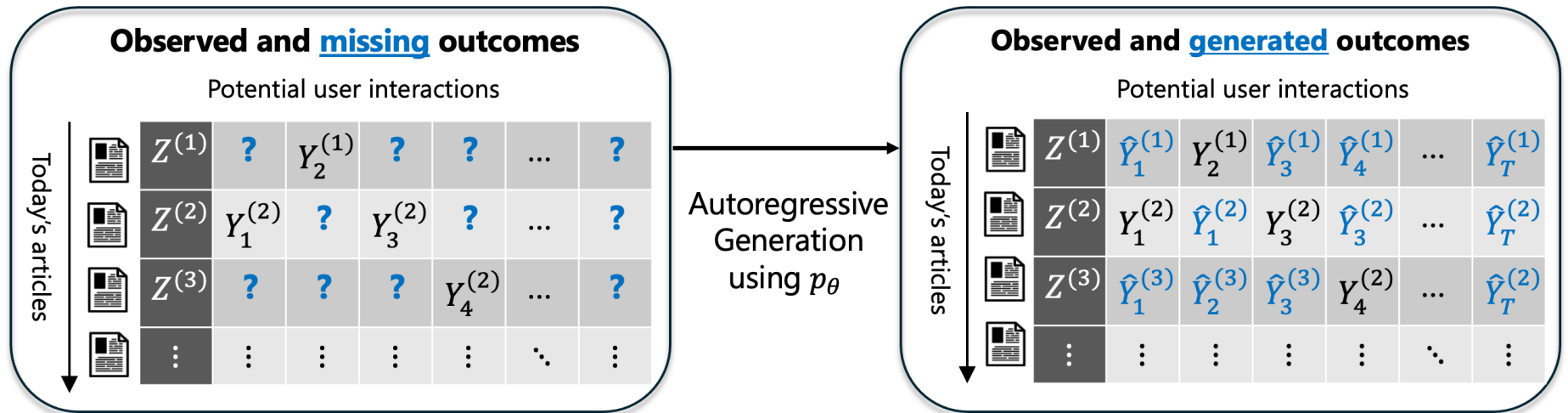
4) Observe outcome



Fill in the table

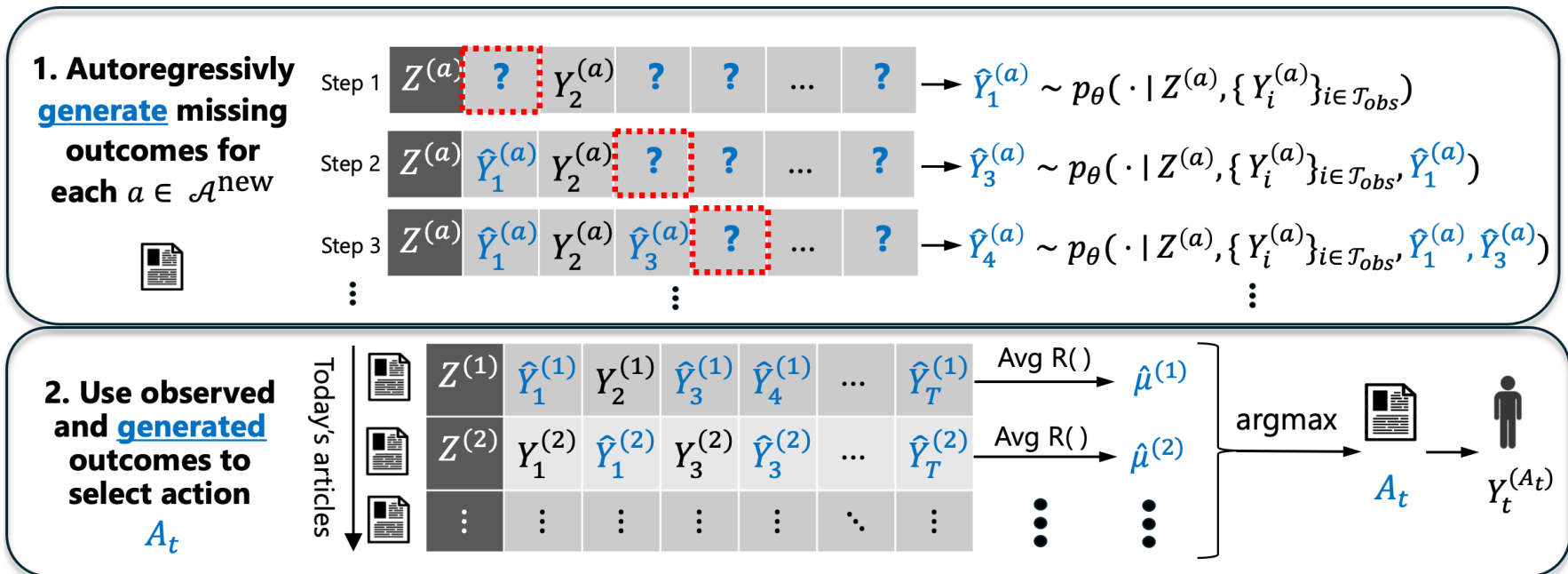
Autoregressive generation

= draw U from a posterior, then generate imagined data



Posterior sampling via autoregressive generation

Optimal decision under imagined data = doing best under U drawn from a posterior



Average regret controlled by prediction error

Two assumptions: 1) bounded rewards, 2) training length sequence exceeds T

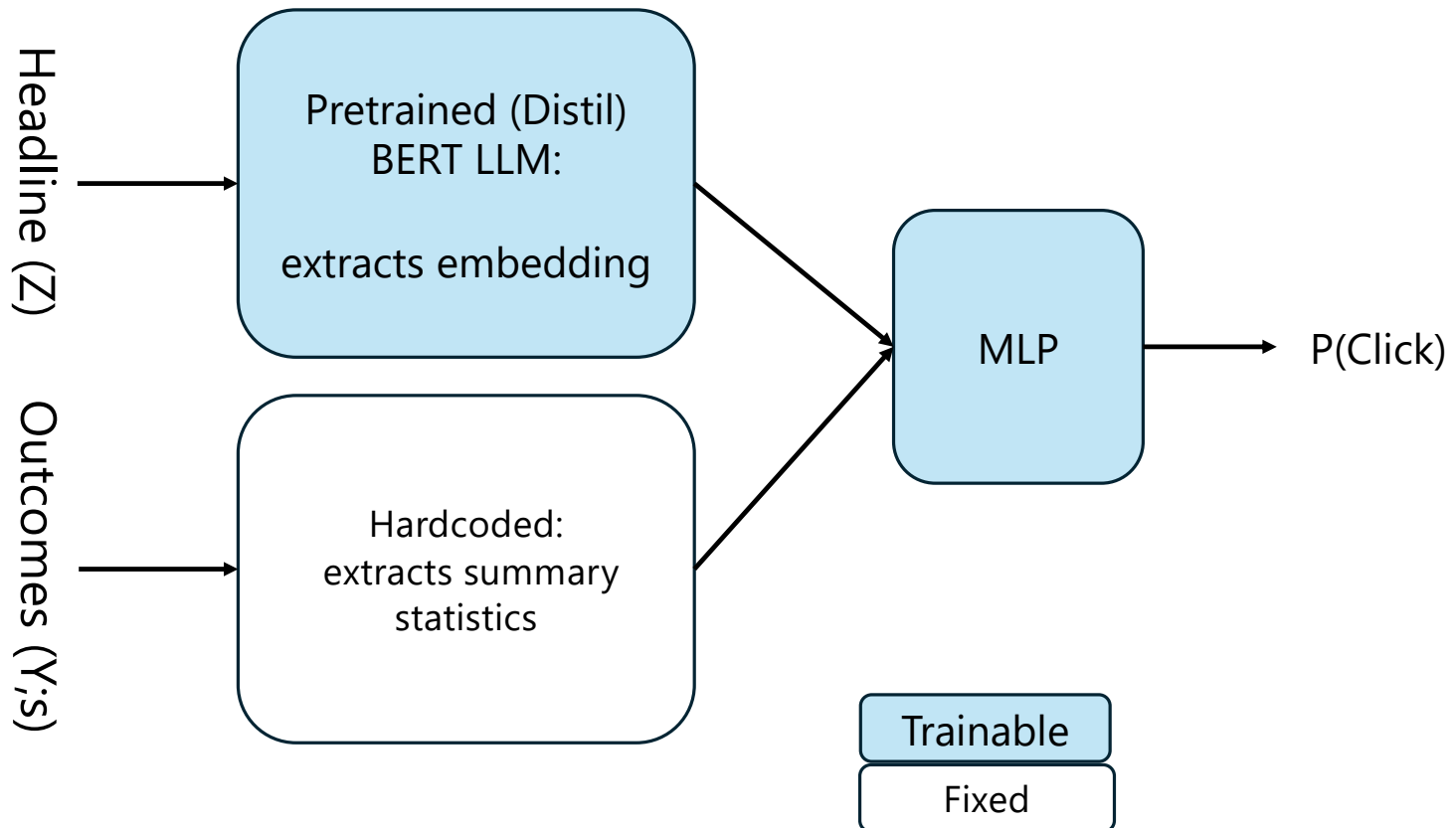
Regret when using autoregressive model p_θ \leq Regret of Thomson sampling from true prior for articles today

$$+ \sqrt{2 \cdot \text{no. articles} \cdot (\ell(p^*) - \ell(p_\theta))}$$

Optimal autoregressive prediction

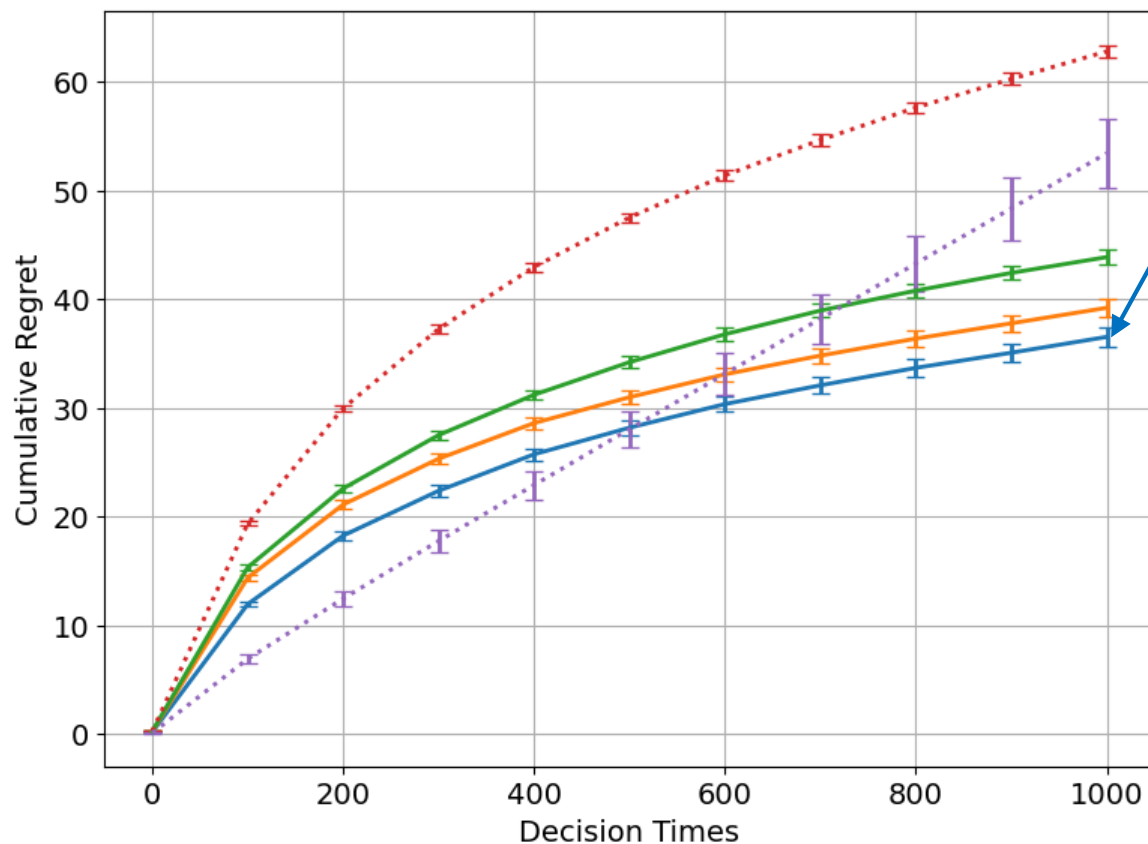
Validation loss

Baby sequence model



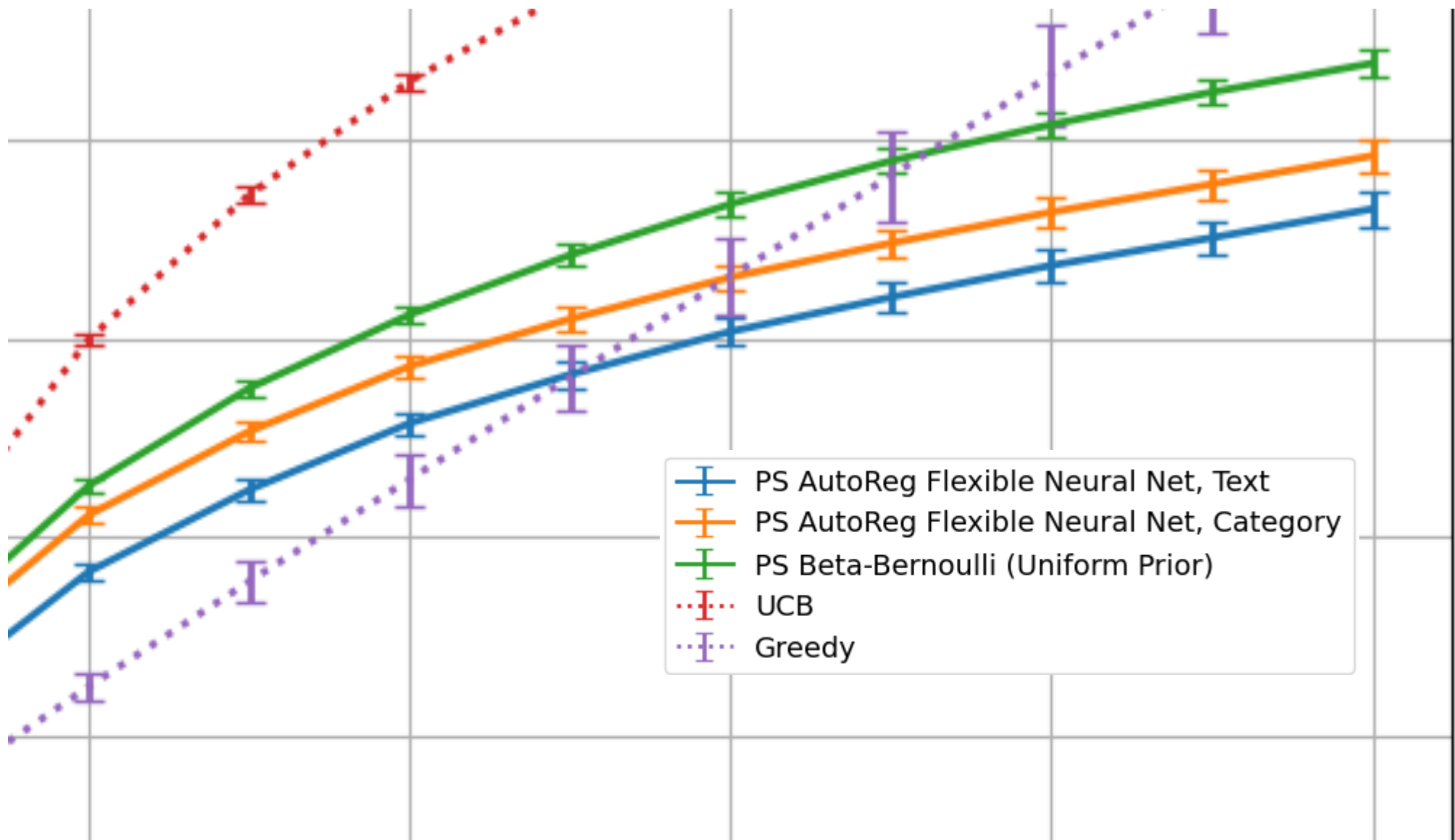
Experiments: Regret

A semi-realistic simulator using public MSN news article data



Autoregressive generation scales to a setting where the best performance requires end-to-end finetuning of an LLM.

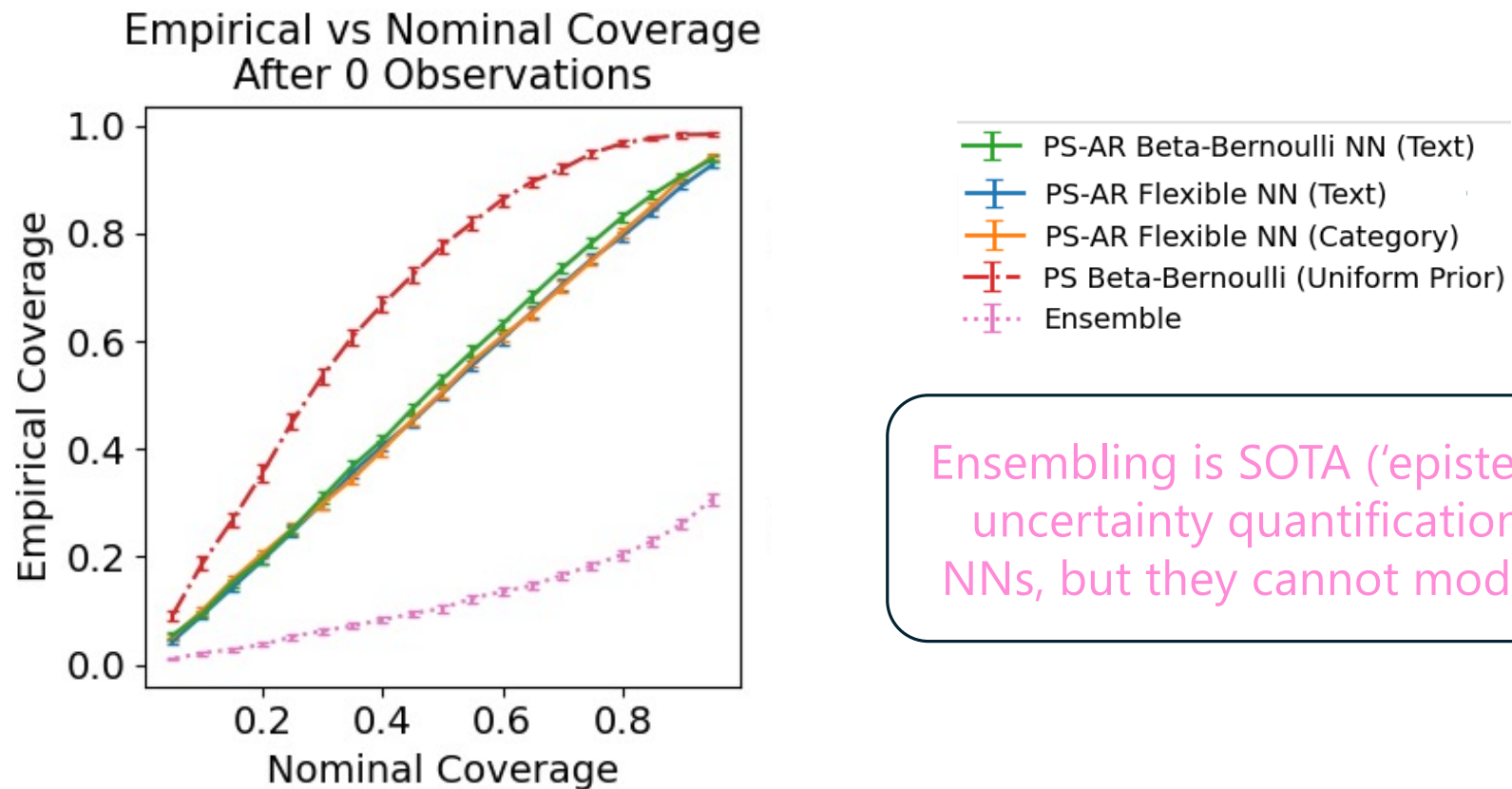
- PS AutoReg Flexible Neural Net, Text
- PS AutoReg Flexible Neural Net, Category
- PS Beta-Bernoulli (Uniform Prior)
- UCB
- Greedy



- PS AutoReg Flexible Neural Net, Text
- PS AutoReg Flexible Neural Net, Category
- PS Beta-Bernoulli (Uniform Prior)
- UCB
- Greedy

Experiments: Coverage

Autoregressive generation mimics proper Bayesian beliefs given headline (text)



Ensembling is SOTA ('epistemic')
uncertainty quantification in
NNs, but they cannot model U.

Summary

<https://arxiv.org/abs/2405.19466>

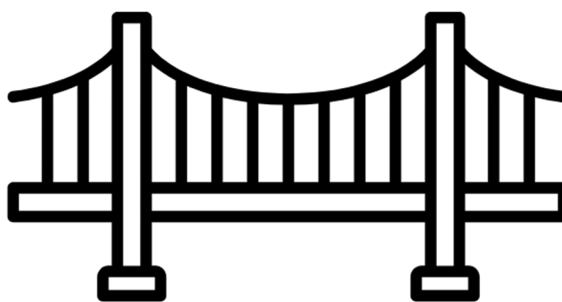
Conceptual: a well motivated problem crystalizing the insights

Algorithmic: link with interactive decision-making

Theory: accurate sequence modeling implies low regret

Experiments: scalable implementations with LLMs.

Generative
Sequence
Modeling



Exploration &
Uncertainty
Quantification