

---

# Adaptive Elicitation of Latent Information Using Natural Language

---

Jimmy Wang<sup>\*1</sup> Thomas Zollo<sup>\*1</sup> Richard Zemel<sup>1</sup> Hongseok Namkoong<sup>1</sup>

## Abstract

Eliciting information to reduce uncertainty about a latent entity is a critical task in many application domains, e.g., assessing individual student learning outcomes, diagnosing underlying diseases, or learning user preferences. Though natural language is a powerful medium for this purpose, large language models (LLMs) and existing fine-tuning algorithms lack mechanisms for strategically gathering information to refine their own understanding of the latent entity. To harness the generalization power and world knowledge of LLMs in developing effective information gathering strategies, we propose an adaptive elicitation framework that *actively* reduces uncertainty on the latent entity. Since probabilistic modeling of an abstract latent entity is difficult, our framework adopts a predictive view of uncertainty, using a meta-learned language model to simulate future observations and enable scalable uncertainty quantification over complex natural language. Through autoregressive forward simulation, our model quantifies how new questions reduce epistemic uncertainty, enabling the development of sophisticated information gathering strategies to choose the most informative next queries. In experiments on the twenty questions game, dynamic opinion polling, and adaptive student assessment, our method consistently outperforms baselines in identifying critical unknowns and improving downstream predictions, illustrating the promise of strategic information gathering in natural language settings.

## 1. Introduction

The performance of many valuable services and systems depends on the ability to efficiently elicit information and

reduce uncertainty about a new environment or problem instance. For example, before an optimal lesson plan can be prepared for a particular student, information must first be gathered about their underlying skills and abilities. Similarly, a patient’s health status must be quickly assessed upon intake, while an online service seeking retention aims to gain a fast understanding of a new customer’s preferences.

Notably, in these (and many other) cases, the object of interest is *latent*, meaning it cannot be directly measured or observed but can only be queried indirectly. This makes gathering information particularly challenging, as it requires carefully designed strategies to infer the latent entity’s characteristics through indirect signals. To achieve efficiency, these strategies must be *adaptive*, dynamically tailoring subsequent queries based on the information gained so far. In the context of student assessment, an adaptive approach might start with broad math questions covering multiple skills. If the student gets a question wrong, the system would then drill down into each relevant skill individually, asking questions of varying difficulty to determine the limits of their proficiency. By progressively refining its queries in this way, the system efficiently maps out the student’s knowledge boundaries and thus reduces uncertainty about their individual skill profile (see Figure 6).

As natural language is a particularly powerful and flexible medium for eliciting such latent information, one might assume that modern large language models (LLMs) (Brown et al., 2020; Bai et al., 2022; DeepSeek-AI et al., 2025) could be helpful in such dynamic information gathering efforts. However, LLMs and existing fine-tuning algorithms often treat uncertainty passively, and lack mechanisms for strategically gathering information to refine their own understanding of the latent entity. While existing LLMs are often trained to instill as much *static* world knowledge as possible (Hendrycks et al., 2021), this world knowledge cannot directly be used to reduce uncertainty about *new, unseen* individuals that the model has little information about.

To harness the generalization power and world knowledge of LLMs to address the renewed uncertainty that arises whenever a new environment or individual is encountered, we introduce an *adaptive elicitation framework* that uses natural language to *actively* reduce uncertainty by simulating future responses. Crucially, our approach leverages *meta-learning*,

---

<sup>1</sup>Columbia University. Correspondence to: Jimmy Wang <jw4209@columbia.edu>, Thomas Zollo <tpz2105@columbia.edu>.

whereby the model is trained on diverse historical question–answer trajectories spanning many latent entities. This meta-trained foundation enables the model to handle new, unseen entities—such as a brand-new student—whose responses are initially unknown and thus create fresh epistemic uncertainty. By aligning a language model’s perplexity objective with the goal of predicting all possible (yet unobserved) answers to the questions we might ask, we transform the challenge of directly modeling a latent entity into a simpler and more scalable problem of predicting *masked future observations* (Ye et al., 2024; Fong et al., 2023). As the model observes each new answer from the individual, it systematically *sharpens its beliefs*, distinguishing between uncertainty it can reduce with further data (epistemic) and the inherent noisiness or variability that remains (aleatoric). Our framework enables a wide range of exciting and impactful applications, e.g., constructing a dynamic diagnostic questionnaire that maximizes the information gained about a patient’s health or generating a personalized set of test questions that yield the most insight into a student’s learning needs (see Figure 1).

**Contribution** In the remainder of this paper, we introduce our novel framework for latent uncertainty reduction using natural language and demonstrate its effectiveness across several key applications. Our work contributes a key conceptual and algorithmic insight to the accelerating field of LLMs: by obviating the need for directly modeling a distribution over the latent entity and instead employing a predictive view of uncertainty, we enable the development of adaptive information gathering strategies that naturally scale with LLM performance, improving as models become more capable. Our adaptive elicitation framework can be applied directly on top of existing LLMs, enabling the use of internet-scale linguistic knowledge to comprehend uncertainty. Through experiments on tasks such as dynamic opinion polling and adaptive student assessments, we illustrate the versatility and significant potential of our framework to enable more efficient and targeted information elicitation in critical domains and applications. Overall, we aim to lay the foundation for future research into rigorous uncertainty quantification and adaptive decision-making using LLMs, highlighting the promise of active, context-aware strategies in solving real-world problems.

## 2. Adaptive Elicitation Framework

In this section, we present an approach to uncertainty quantification and adaptive question selection in natural language settings where the latent entity cannot be directly modeled. We adopt a *predictive view* of uncertainty: rather than specifying a direct prior or complete model of the latent entity, we focus on how well the model can predict and quantify uncertainty over future observations of that entity. Our method:

(1) **Meta-learns** a predictive language model from historical question–answer data. (2) Uses this model to **quantify uncertainty** about future or unobserved answers from *new, unseen* latent entities, using autoregressive forward simulation to efficiently distinguish between *epistemic* and *aleatoric* uncertainty. (3) Accurately quantifies and sharpens beliefs given new information, and dynamically **adapts question selection** to elicit information that optimally reduces uncertainty. A key advantage of our method is that we can apply our meta-learning method directly on existing pre-trained LLMs, augmenting uncertainty quantification with internet-scale world knowledge.

### 2.1. Problem Formulation

We consider an unobservable latent entity  $U \in \mathcal{U}$  (e.g., a student’s skill profile or a patient’s health status). We query  $U$  by posing a question  $X \in \mathcal{X}$  (in natural language) and observing an answer  $Y \in \mathcal{Y}$  drawn from

$$\text{Answer } Y \sim Q(\cdot | \text{Question } X, \text{Latent } U), \quad (1)$$

where  $Q$  is the ground truth distribution. Our two primary goals are to: (1) *Quantify* our uncertainty about  $U$  based on observed question–answer pairs. (2) *Reduce* that uncertainty by adaptively choosing which questions  $X$  to ask next.

**Pre-Training** We assume that we have an abundance of historical data, where a model can learn from *past trajectories* to inform adaptive elicitation about *new, unseen* individuals from which we wish to gather data. We have a collection of entities  $U \in \mathcal{U}_{\text{train}}$ . For each latent entity  $U$  we have access to a sequence of questions and answers  $(X_{1:T}^{(U)}, Y_{1:T}^{(U)})$  produced by it. Then our historical pre-training data consists of:  $\mathcal{D}_{\text{train}} := \{X_{1:T}^{(U)}, Y_{1:T}^{(U)} : U \in \mathcal{U}_{\text{train}}\}$ . For example, an online tutoring service may have an abundance of data about previous students that they may utilize.

**Test-Time Adaptive Selection** After pre-training, we wish to quantify and reduce uncertainty about new, unseen, latent entities  $U_{\text{new}}$  (e.g. a new student or patient). For each  $U_{\text{new}}$  we have  $T$  rounds where we can sequentially ask questions and receive responses. At each  $t = 0, 1, \dots, T - 1$ , the model can adaptively choose a question  $X_{t+1} \in \mathcal{X}_{t+1}$  based on previous feedback  $\mathcal{H}_t := (X_{1:t}^{U_{\text{new}}}, Y_{1:t}^{U_{\text{new}}})$  and receives an answer  $Y_{t+1} \in \mathcal{Y}_{t+1}$ .

We will evaluate the model on its ability to predict unobserved answers  $Y_{T+1:\infty}$  generated by the latent entity to any future questions  $X_{T+1:\infty}$ . For example, we may conduct a survey where we can only include 5 questions, but we wish to know the answer to 1000 additional questions. Being able to predict  $Y_{T+1:\infty}$  requires the ability to adaptively collect relevant information  $Y_{1:T}$ , which in turn requires the ability to both quantify and reduce remaining uncertainty about

the entity  $U_{\text{new}}$ . We detail our approach in the following sections.

## 2.2. Quantifying Uncertainty Using a Predictive Model

Traditional approaches to model uncertainty may try to model  $U$  directly (e.g. by assigning a probability distribution over a structured latent space) (Blei et al., 2003; Salakhutdinov & Mnih, 2007). However, specifying such models for complex human-generated responses can be both restrictive and infeasible. For example, it is unclear how to define an explicit parametric model to represent an individual’s political opinions. Instead, we adopt a *predictive view* of uncertainty that avoids the need to directly model latent variables. This approach allows us to train autoregressive models directly in the space of natural language, enabling flexible and scalable modeling of its full complexity.

Define the *epistemic uncertainty* to be uncertainty that can be reduced with more information, and the *aleatoric uncertainty* to be uncertainty due to random noise or variation that cannot be reduced by observing more data. Our key observation is that if we were to observe infinite data  $Y_{1:\infty}$  produced by the entity, all *epistemic* uncertainty about the entity would disappear. Intuitively, if a teacher could observe a student’s answers to a very large set of questions, that teacher could probably completely predict a student’s future answers with errors only due to *aleatoric* variation. Examples of this aleatoric uncertainty could include random noise or intrinsic uncertainty in the student’s own decision process. This idea is in line with classical views that treat latent variables as unobserved data (Rubin, 1976; Lindley, 1965; Hill, 1968; Dawid, 1984), as well as more modern treatments (Fong et al., 2023; Ye et al., 2024; Zhang et al., 2024).

Under this view, epistemic uncertainty is naturally the uncertainty due to *missing data*: specifically, the uncertainty about unobserved future responses  $Y_{t+1:\infty}$  given the current information  $Y_{1:t}$ . Thus, our objective is to provide accurate uncertainty estimates over missing data  $Y_{t+1:\infty}$ . In order to quantify uncertainty about  $Y_{t+1:\infty}$ , we first build off the notion of *entropy*. Given a distribution  $P$  with density  $p(\cdot)$ , the entropy and conditional entropy over an answer  $Y \in \mathcal{Y}$  are defined as  $H_P(Y) = \sum_{y \in \mathcal{Y}} p(y) \log p(y)$ ,  $H_P(Y \mid \cdot) = \sum_{y \in \mathcal{Y}} p(y \mid \cdot) \log p(y \mid \cdot)$ .

Next, let

$$P(Y_{t+1} = y \mid X_{1:t}, Y_{1:t}, X_{t+1} = x) = p_t(y \mid X_{1:t}, Y_{1:t}),$$

which also induces the conditional entropy  $H_P(Y_{t+1} \mid X_{1:t}, Y_{1:t})$ . Using this notation, we are interested in measuring the uncertainty over missing data

$$\begin{aligned} \text{Uncertainty (Future Answers \mid Current Info)} = \\ H_P(Y_{t+1:\infty} \mid X_{1:t}, Y_{1:t}). \end{aligned} \quad (2)$$

## Example: Adaptive Student Assessment

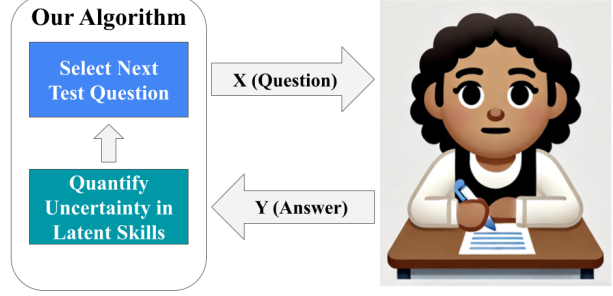


Figure 1. Our algorithm can adaptively elicit information from a latent entity via natural language interaction. For example, in assessing a new student, the system may ask questions in areas where the student’s abilities are not yet known, to maximize the information gained from each question and efficiently reduce uncertainty about the student’s individual skill profile.

Once we have this estimate, we can adaptively choose questions that reduce the greatest amount of uncertainty about the missing data  $Y_{t+1:\infty}$  (see Section 2.4 for more details). Notice that this approach to uncertainty quantification works directly in the space of observables  $X_{1:\infty}, Y_{1:\infty}$ , and does not require any explicit modeling of the latent  $U$ . If we have a predictive distribution over future answers  $Y_{t+1:\infty}$  given previous observations  $X_{1:t}, Y_{1:t}$  for every  $t$ ,

$$P(Y_{t:\infty} \mid X_{1:t}, Y_{1:t}) = P(\text{Future Answers} \mid \text{Current Info}),$$

then we can exactly calculate the entropy term in Equation (2). Thus, we directly train an autoregressive predictive model  $p_\theta(Y_{t+1:\infty} \mid X_{1:t}, Y_{1:t})$  by optimizing the parameters  $\theta \in \Theta$  in order to model these quantities.

## 2.3. Meta-Learning an Autoregressive Predictive Model

To obtain the most accurate estimates of uncertainty, the ideal strategy would be to use the ground-truth answer distribution  $Q$  in Equation (1) as the predictive distribution over future observations. In particular, we aim to approximate the conditional distribution  $q(Y_{t+1:\infty} \mid X_{1:t}, Y_{1:t})$  induced by  $Q$ , which represents the true distribution over future answers given past interactions. Under this setup, the corresponding conditional entropy  $H_Q(Y_{t+1:\infty} \mid X_{1:t}, Y_{1:t})$  would yield exact measures of uncertainty. Since  $Q$  is unknown in practice, our goal becomes to approximate the conditional  $Q$  as closely as possible by training a model  $p_\theta$  in order to produce reliable uncertainty estimates.

To approximate  $Q$ , we assume access to historical data from a collection of latent entities  $\mathcal{U}_{\text{train}}$ . Each entity  $U \in \mathcal{U}_{\text{train}}$  is associated with a sequence of question–answer pairs  $\{(X_{1:T}^{(U)}, Y_{1:T}^{(U)})\}$ , where  $Y \sim Q(\cdot \mid X, U)$  and  $U$  is

sampled from a prior distribution. Our first step is to meta-train an autoregressive language model  $p_\theta$  on this historical data consisting of sequences of questions and answers from various latent entities

$$\mathcal{D}_{\text{train}} := \{X_{1:T}^{(U)}, Y_{1:T}^{(U)} : U \in \mathcal{U}_{\text{train}}\}.$$

In the student assessment example,  $\mathcal{D}_{\text{train}}$  may be a historical dataset of past students, each with an associated sets of test questions and answers. For simplicity, we assume that each sequence is of length  $T$ , but our framework is agnostic to differing sequence lengths.

**Objective** Define a sequence of previous observations  $\mathcal{H}_t := \{X_{1:t}, Y_{1:t}\}$ . We train our autoregressive language model  $p_\theta$ , parameterized by  $\theta \in \Theta$ , to output one-step probabilities over future answers conditioned on previous observations (i.e., question/answer pairs)  $p_\theta(Y_{t+1} | \mathcal{H}_t, X_{t+1} = x)$ , inducing a joint distribution over future outcomes

$$p_\theta(Y_{t+1:\infty} | \mathcal{H}_t, X_{t+1} = x_{t+1}, X_{t+2} = x_{t+2}, \dots) = \prod_{s=t+1}^{\infty} p_\theta(Y_s | \mathcal{H}_{s-1}, X_s = x_s). \quad (3)$$

The training objective for our model is then to optimize the joint log likelihood/marginal likelihood of the observed sequence within the historical dataset

$$\max_{\theta \in \Theta} \left\{ \frac{1}{|\mathcal{U}_{\text{train}}|} \sum_{U \in \mathcal{U}} \sum_{t=1}^T \log p_\theta(Y_t^U | \mathcal{H}_{t-1}, X_t^U = x_t) \right\}. \quad (4)$$

After optimizing our model  $p_\theta$ , we can now use it to approximate the uncertainty estimates  $H_Q(Y_{t+1:\infty} | X_{1:t}, Y_{1:t}) \approx H_{p_\theta}(Y_{t+1:\infty} | X_{1:t}, Y_{1:t})$ . To show that optimizing this objective is optimal for approximating  $Q$ , we first note that maximizing the objective in Equation (4) is equivalent to optimizing an empirical version of the cross entropy  $\mathbb{E}_Q[\log p_\theta(Y_{1:T} | X_{1:T})]$ . Expanding this loss, we can see that

$$\begin{aligned} \max_{\theta} \mathbb{E}_Q[\log p_\theta(Y_{1:T} | X_{1:T})] &= \\ \max_{\theta} \{\mathbb{E}_Q[\log q(Y_{1:T} | X_{1:T})] - \text{D}_{\text{KL}}(Q \| p_\theta)\} &= \\ \mathbb{E}_Q[\log q(Y_{1:T} | X_{1:T})], \end{aligned}$$

where  $\text{D}_{\text{KL}}(Q \| Q) = 0$ . An equivalent interpretation of maximizing the joint log likelihood is that we are minimizing the KL divergence between  $p_\theta$  and  $Q$ , leading to accurate downstream uncertainty estimates. By optimizing this objective over historical data, our aim is for the model to learn meta-learn structures and patterns that will be useful for adaptive testing over *new*, *unseen* entities.

**Training.** For training, we process each sequence of questions and answers  $\{X_{1:N}^{(U)}, Y_{1:N}^{(U)}\}$  corresponding to a latent entity  $U$  by sequentially arranging them into one long natural language string  $(X_1^{(U)}, Y_1^{(U)}, X_2^{(U)}, Y_2^{(U)}, \dots)$ . We assume that the probability of a response to a question is independent of the ordering of the earlier questions and answers, and we randomly permute the order of the question/answer pairs within each entity's sequence during training. Then we optimize a language model to predict each answer  $Y_t$  conditioned on the current question  $X_t$  and previous observations  $\mathcal{H}_{t-1}$ . To do so, we apply a gradient mask that masks out tokens which do not correspond to any  $Y_i$ . We use stochastic gradient descent procedures to optimize the training loss.

## 2.4. Adaptive Question Selection by Future Simulation

Having trained a predictive model  $p_\theta$  from historical data, we can use this model to quantify uncertainty about future observations generated by the latent entity, and take actions to reduce said uncertainty. Given a new latent entity  $U_{\text{new}}$ , we may also be interested in different targets of uncertainty: for example, the full sequence of future answers ( $Z = Y_{1:\infty}$ ), a specific subset of questions, or the answer to a particular question ( $Z = Y$ ). To make our notation more general, we notate  $Z$  as the object we wish to understand. Our goal is to reduce uncertainty about  $Z$  by sequentially choosing questions  $X$  that are most informative — i.e., those that will optimally reduce the model's uncertainty about  $Z$  (See Figure 6). For example a tutor may be interested in understanding which questions will reveal the most about the student's understanding.

**Setup** As described in Section 2.1, we operate in an adaptive setting at test time. At each round  $t = 0, 1, \dots, T - 1$ , we may select a question  $X_{t+1} \in \mathcal{X}_{t+1}$  based on the interaction history so far,  $\mathcal{H}_t := (X_{1:t}^{U_{\text{new}}}, Y_{1:t}^{U_{\text{new}}})$ , and receive an answer  $Y_{t+1} \in \mathcal{Y}_{t+1}$ .

To quantify informativeness, we define the information gain from a question-answer pair  $(X_{t+1}, Y_{t+1})$  as:

$$\begin{aligned} \text{IG}_t(Z; (X_{t+1}, Y_{t+1})) &= \\ H(Z | \mathcal{H}_t) - H(Z | \mathcal{H}_t \cup (X_{t+1}, Y_{t+1})). \end{aligned} \quad (5)$$

This measure quantifies the reduction in entropy about  $Z$  after observing a new interaction, by quantifying the difference between the current uncertainty  $H_{p_\theta}(Z | \mathcal{H}_t)$  and the uncertainty after observing  $(X_{t+1}, Y_{t+1})$ ,  $H(Z | \mathcal{H}_t \cup (X_{t+1}, Y_{t+1}))$ . Since we do not yet know  $Y_{t+1}$  when choosing  $x_{t+1}$ , we can instead quantify the *expected* reduction in uncertainty by simulating potential answers using our meta-learned model  $p_\theta$ . This idea leads to the *Expected*



*Information Gain (EIG)* (Chaloner & Verdinelli, 1995):

$$\text{EIG}_t(Z; x_{t+1}) = H_{p_\theta}(Z | \mathcal{H}_t) - \mathbb{E}[H_{p_\theta}(Z | \mathcal{H}_t \cup (x_{t+1}, Y_{t+1}))], \quad (6)$$

where we use our meta-learned model  $p_\theta(\cdot)$  to simulate  $Z$  and  $Y_{t+1} \sim p_\theta(\cdot | \mathcal{H}_t, X_{t+1} = x_{t+1})$  in the expectation. To calculate the EIG for multiple choices of  $x$ , we have

$$\text{EIG}_t(Z; (x_{t+1}, x_{t+2}, \dots, x_{t+K})) = H_{p_\theta}(Z | \mathcal{H}_t) - \mathbb{E}[H_{p_\theta}(Z | \mathcal{H}_t \cup \bigcup_{i=t+1}^{t+K} (x_i, Y_i))], \quad (7)$$

where we autoregressively simulate  $Y_{t+1:t+K}$  from our meta-learned model. This quantity naturally quantifies the amount of epistemic uncertainty we expect to reduce by choosing a set of questions. If the EIG is very small, then this implies that the reduction in entropy is small and therefore this information is not informative. This could be because there is a lot of aleatoric uncertainty, such that the information gathered is very noisy, or due to the fact that the information gathered is not relevant to the object of interest.

**Question Selection Policies** To select the optimal question at each time  $t$ , we would ideally like to choose a question selecting policy  $\pi : \mathcal{H} \mapsto \mathcal{X}$ , where  $X_{t+1} \sim \pi(\cdot | \mathcal{H}_t)$ , such that we maximize

$$\arg\max_{\pi} \mathbb{E}_{X \sim \pi(\cdot)} [\text{EIG}_t(Z; X_{t+1:T})], \quad (8)$$

where  $\text{EIG}_t(Z; X_{t+1:T})$  is defined in Equation (7). To approximate this quantity, we use our meta-learned model  $p_\theta$  to autoregressively simulate possible future answers  $Y_{t+1:T} \sim p_\theta$ , and the chosen policy  $\pi$  to simulate question choices  $X_{t+1:T} \sim \pi(\cdot)$ . Through autoregressive future simulation, we can optimize for our question selection policy  $\pi$ . While it is possible to calculate the discrete optimal  $x_{t:T}$  that maximizes this objective, it can be intractable as simulating  $X_{t+1:T}, Y_{t+1:T}$  is combinatorial in the number of steps. Instead, we introduce two procedures that show strong practical performance while having feasible computational cost.

**Greedy Selection.** A simple question selection policy  $\pi^{\text{greedy}}$  is to first enumerate the candidate questions  $x \in \{x_1, \dots, x_k\}$ . Then for each  $x_j$ , calculate the one-step expected information gain  $\text{EIG}_{t:t+1}(Z; x_j)$ . Finally, choose the  $x_j$  that maximizes this quantity. Concretely,

$$\pi_t^{\text{greedy}} := \arg\max_x \text{EIG}_{t:t+1}(Z; x).$$

Although greedy, this policy often performs well in practice and is computationally simpler than globally optimal planning. We establish the theoretical validity of this procedure

in Section B, where Proposition B.3 bounds the performance gap between a full combinatorial planning approach and the greedy selection procedure.

**Lookahead / Monte Carlo Planning.** To account for multi-step effects (e.g., a question that might not immediately reduce much uncertainty but paves the way for more informative follow-ups) and to better approximate the combinatorial quantity in Equation (8), we can apply standard *Monte Carlo Tree Search* (MCTS) techniques from reinforcement learning (Browne et al., 2012; Silver et al., 2016). We use a simple instantiation with strong empirical performance and leave more complex variants to future exploration.

With an MCTS policy  $\pi^{\text{MCTS}}$ , we sample entire simulated question-answer sequences using the meta-learned model  $p_\theta$  up to depth  $d$  to estimate the cumulative information gain. In order to simulate future responses, we use  $\pi^{\text{greedy}}$  to select questions and the Information Gain (5) as proxy rewards. Starting at time  $t$ , we first calculate  $\text{EIG}_{t:\infty}(Z; X_{t+1:\infty})$  for each  $x \in \mathcal{X}_t$ , and choose the top  $K$  questions. For each of the  $K$  questions, we then simulate  $N$  futures up to depth  $d$ . For each sample path  $i \in [N]$ , we receive reward  $r^{(i)}(x) = \text{IG}_t(Z; (X_{t+1:t+d}^{(i)}, Y_{t+1:t+d}^{(i)}))$ , where questions are sequentially selected using  $\pi^{\text{greedy}}$  and answers are simulated using the meta-learned model  $p_\theta$ . Finally, the MCTS policy chooses an action as

$$\pi^{\text{MCTS}} := \arg\max_{x \in \mathcal{X}_t} \frac{1}{N} \sum_{i=1}^N \text{IG}_t(Z; (X_{t+1:t+d}^{(i)}, Y_{t+1:t+d}^{(i)})).$$

Though more expensive computationally, we find that  $\pi^{\text{MCTS}}$  can find better long-horizon query strategies.

### 3. Experiments

To rigorously benchmark adaptive information gathering strategies for LLMs, we require datasets that (i) capture diverse latent entities or hidden factors, (ii) provide many possible queries about these entities, and (iii) for each entity, link some queries to corresponding ground-truth answers. Such an experimental setup allows us to assess the ability of an LLM and/or particular algorithm to strategically select questions in order to reduce uncertainty about the latent entity. Ideally, each dataset reflects the real-world complexities of human-generated responses, while still providing enough structure for robust evaluation of different query selection policies. In practice, a large pool of possible questions with many ground truth answers is essential, since it allows an adaptive strategy to actively and deeply explore the latent entity along many dimensions, while still leaving unobserved data for evaluation. Then, each latent entity (e.g., a survey respondent’s political stance, a stu-

dent’s hidden skill profile, or the identity of a secret object) can be progressively unveiled by observing how it answers newly selected queries. Such design criteria enable controlled, quantitative evaluations of LLMs under interactive, information gathering scenarios.

Our experiments focus on three applications: the “Twenty Questions” game (using our novel and publicly available dataset, described below), opinion polling, and student assessment. In each scenario, the objective is to adaptively select questions that reveal as much information as possible with respect to a separate (though potentially overlapping) set of target questions. Questions are chosen one at a time, and each new question–answer pair is appended to the LLM’s context before proceeding. For every experiment, we start with a dataset containing a collection of latent entities  $U$ , each associated with a set of questions  $X^U$  and answers  $Y^U$ . To train our meta-learned model  $p_\theta$ , we split each dataset by groups of latent entities into training, validation, and test sets. We first meta-learn  $p_\theta$  on question–answer pairs corresponding to the training entities, after which we evaluate how effectively the model quantifies and reduces epistemic uncertainty about observations generated by test entities. Further details regarding the datasets, training procedure, baseline comparisons, and evaluation metrics are provided below.

### 3.1. Twenty Questions Dataset

The classic “Twenty Questions” game epitomizes our core goal of reducing uncertainty about a hidden entity through targeted queries. Specifically, the object (such as an animal, a musical instrument, or any of a wide range of other concepts) serves as the latent factor  $U$  that cannot be directly observed. A player (or model) must identify  $U$  by posing a sequence of questions and observing the corresponding answers (e.g., “no,” “maybe,” or “yes”). Hence, the game inherently captures the essence of characterizing a latent entity by uncovering how it generates answer to different possible questions. To operationalize this game for benchmarking, we construct a novel “Twenty Questions” dataset from a curated set of objects in the THINGS database (Hebart et al., 2019), each serving as a potential hidden entity. For each object, we produce a diverse set of candidate questions (e.g., “Does it have four legs?”, “Is it edible?”, “Is it used for entertainment?”) together with the corresponding answers, generated by a top-quality LLM (Claude-3.5-Sonnet). In total, the dataset contains 800 objects, each with answers to a set of 1200 questions. By treating each object as a distinct latent entity, we capture a broad spectrum of scenarios, ranging from everyday items (“banana,” “telephone”) to more uncommon concepts (“violin,” “canoe”). We note that the absolute correctness of Claude’s answers is not crucial, as our goal is for our model to learn the underlying data-generating process governing which answers appear,

given specific questions and objects. Our dataset is publicly available,<sup>1</sup> including the complete set of objects, curated questions, generated answers, and relevant metadata.

### 3.2. Other Datasets

**OpinionQA (Santurkar et al., 2023)** Originally created to evaluate the alignment of LLM opinions to those of 60 US demographic groups, this dataset contains 1498 multiple choice political questions answered by a diverse collection of survey respondents. These questions target various political issues ranging from abortion to automation. For each question  $X$ , the multiple choice answer corresponds to the observable feedback  $Y$ , and the survey respondent’s latent political preference corresponds to the unobservable  $U$ .

**EEDI Tutoring Dataset (Wang et al., 2020)** EEDI is an online educational and tutoring platform that serves millions of students around the globe. This dataset includes a collection of 938 math questions focusing on various areas such as algebra, number theory, and geometry, as well as individual responses from many students. Each question is a multiple choice question with four answers that includes a visual diagram as well as associated text. The student’s true mathematical ability  $U$  generates the student’s answer  $Y$  to the math question  $X$ .

### 3.3. Meta-Training Details

We first split the training datasets by entity into train, validation, and test with a 70%, 15%, 15% split. To meta-train our model, we initialize a pre-trained Llama-3.1-8B model in FP16 precision and use LoRA (Hu et al., 2021) to finetune our model with parameters  $\alpha = 24$ , rank= 8, and dropout= 0.1. Additional details are shown in Appendix C.1

### 3.4. Baselines

Here we describe the baselines to which we compare our algorithm; each consists of an approach to model fine-tuning, and an approach to question selection. As in our method, each chosen question–answer pair is appended to the LLM context for predicting unseen answers.

**Base LLM** First, we consider a simple baseline. For an LLM we use Llama-3.1-8B, from which our meta-trained model is initialized, with no additional fine-tuning; question selection is performed randomly.

**In-Context Tuning (ICT)** Next, we consider a typical in-context learning (ICL) baseline. First, we meta-train the

<sup>1</sup><https://huggingface.co/datasets/namkoong-lab/TwentyQuestions>

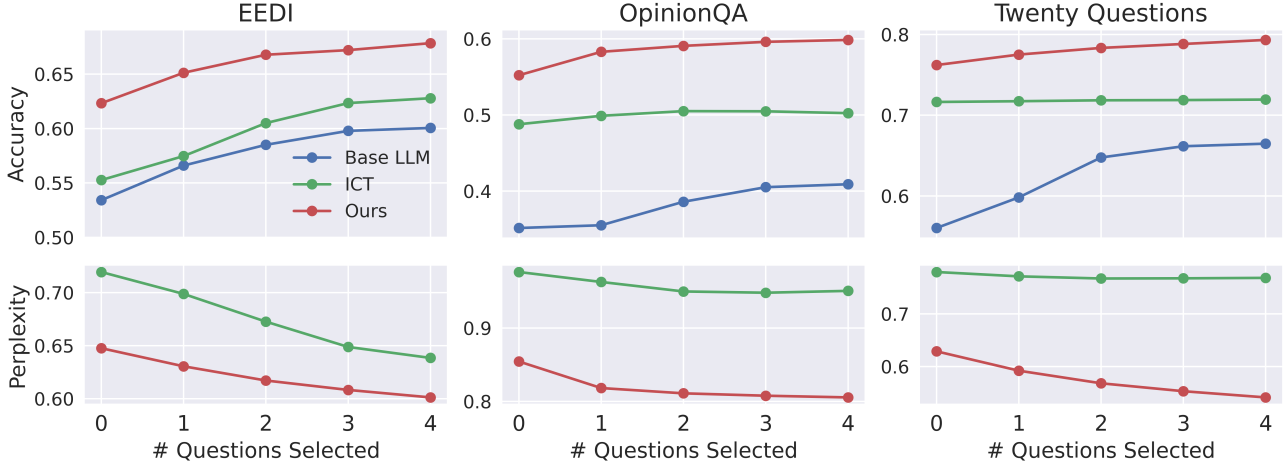


Figure 2. Accuracy (top) and perplexity (bottom) of our adaptive elicitation framework compared to baseline methods across three datasets: OpinionQA, EEDI student assessment, and Twenty Questions. The x-axis represents the number of questions selected. Our method works best to gather information and accurately characterize the latent in each case. Each plot is the average of 10,000 simulations across unseen entities.

model via In-Context-Tuning (Chen et al., 2022), where the objective is to predict the label for a query example given some number of in-context support examples. Then, questions are selected based on embedding similarity to the target questions that we aim to answer (Liu et al., 2021). We use the same model and parameters as described in Section 3.3, and we use Alibaba-NLP/gte-large-en-v1.5 as our embedding model.

### 3.5. Evaluation

To evaluate how well each method can ask targeted questions to reduce uncertainty about the latent entity, we perform 10,000 trials, where on each trial we randomly select an entity and apply our algorithm (and baselines). For each trial and its corresponding entity, we randomly select a pool of  $N$  **candidate questions** from which the methods can sequentially choose questions to ask, and randomly select  $K$  held-out **target questions**. The objective is to sequentially choose optimal questions from the candidate questions to reduce the most uncertainty about the held-out target questions for the entity. In our experiments, we choose  $N = 20$  and  $K = 5$ , but we include ablations that vary these quantities in Section 4.5. We evaluate the performance on the target questions with four metrics: (1) Accuracy, (2) Perplexity (Jelinek et al., 1977), (3) Expected Calibration Error (Guo et al., 2017), and (4) Brier Score (Brier, 1950).

## 4. Results and Discussion

In this section, we empirically study the following questions: (1) Can our framework be used to adaptively select questions to reduce uncertainty and elicit information about the

latent entity? (2) Do we generate reasonable posterior probability updates and reduce uncertainty as more information is gathered? (3) When is this adaptive procedure particularly helpful, and when is advanced planning (i.e., MCTS) most important? (4) How crucial is our training procedure for producing actionable uncertainty quantification? (5) Does the performance of our framework improve with a better underlying LLM? Throughout, we connect these findings to the paper’s broader motivation: the importance of adaptive strategies to eliciting information efficiently in real-world scenarios.

### 4.1. Overall Gains from Adaptive Elicitation

Overall results for our method and 2 baselines across all 3 datasets are shown in Figure 2. The top row of plots record accuracy on the target questions, while the bottom row record perplexity (or negative log-likelihood loss). The Base LLM is omitted on bottom for ease of visualization. Our framework is applied using the greedy EIG strategy. In both figures, the X-axis records the number of questions that have been selected so far.

Across all 3 datasets and both metrics, our algorithm most effectively characterizes the latent entity by predicting the answers to target questions (we show similar results for Brier Score in Figure 7). Further, our algorithm consistently improves its characterization as more information is gathered, whereas gathering more questions based on embedding distance does not always help. Overall, our adaptive elicitation framework proves effective in gathering information and reducing uncertainty across 3 diverse domains.

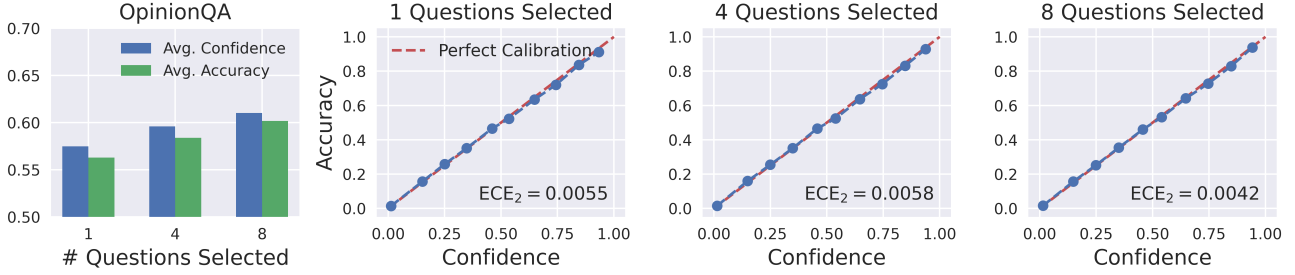


Figure 3. Reliability diagrams comparing confidence and accuracy after different numbers of selected questions (and observed answers). Our model maintains well-calibrated uncertainty estimates, increasing both confidence and accuracy as more questions are asked.

## 4.2. Uncertainty Quantification

A cornerstone of our approach is using *predictive* perplexity as an indicator of uncertainty to guide the adaptive strategy; this makes sense only if our model’s probabilities correctly reflect confidence about unseen data. To assess this, we examine calibration, or the extent to which the model’s confidence reflects its prediction accuracy.

For each dataset, we plot reliability diagrams (Guo et al., 2017) of confidence vs. accuracy, where perfect calibration lies on the  $y = x$  line, and record Expected Calibration Error (ECE). Both the reliability diagram and ECE are produced by separating predictions into 10 bins by confidence, and comparing the average confidence and accuracy for each bin. Results are shown after 1, 4, and 8 questions are selected, and the far left subfigure displays overall average confidence and accuracy for each setting.

Results for OpinionQA are shown in Figure 3, while EEDI and Twenty Questions are shown in Appendix Figures 8 and 9. For all 3 datasets, we observe that the predicted probabilities lie close to the diagonal of perfect calibration—our model’s confidence aligns well with actual accuracy. As more questions are observed, the model’s average confidence (and accuracy) both go up, confirming that uncertainty diminishes in an intuitive way. In the motivating student-assessment scenario, this means that by asking just a few strategically chosen questions, the model not only improves its predictions but also becomes *more certain* in them. For a high-stakes application such as medical diagnostics or skill placement exams, it is crucial to know when a model has enough data to be sure in its predictions, versus when it is still uncertain; these calibration results confirm our framework performs well in this sense.

## 4.3. When is Adaptivity Most Helpful?

Having established that our adaptive question selection method is generally effective at quantifying uncertainty and eliciting information about some latent, we next examine *when* such a procedure is most helpful. In particular, we

hypothesize that adaptive strategies are especially important in characterizing features of the latent which are relatively rare in the population. As a concrete example, while many students may have overlapping weaknesses (e.g., many get the same test question wrong), it can be harder to learn that a particular student is struggling in an area where other students generally do not. An adaptive strategy could help by selecting a test question that most find easy but this student may answer incorrectly.

To investigate this hypothesis, we specify two different subgroups of questions as targets by running an evaluation where for each target question in the subgroup, the entity’s answer must have probability less than either 50% (“medium”) or 30% (“hard”) across the population. We use our meta-trained model with random, EIG, and MCTS question selection, and record results after  $N$  questions have been selected. Results are shown in Figure 4. For each question subgroup (as well as all questions from the previous experiment), we record on the y-axis the relative accuracy gain from using EIG or MCTS, compared to random selection.

First, we notice that the more advanced MCTS planning strategy outperforms EIG in all cases, and both always outperform random. Intuitively, while a greedy strategy picks the single next question that locally maximizes immediate information gain, it may miss questions whose short-term yield seems small but that pave the way for far more revealing follow-up queries. By looking multiple steps ahead, MCTS better accounts for how each query reshapes future options, often enabling it to find more globally optimal questioning strategies. This means that given a good model for uncertainty quantification, we can improve our results by spending more compute, indicating good scaling behavior in our algorithm.

Next, we observe trends across different subgroups of questions. In all 3 example applications, adaptivity and planning have a massive impact on the ability to answer hard questions compared to random question selection. For EEDI and Twenty Questions the percent gain over random with



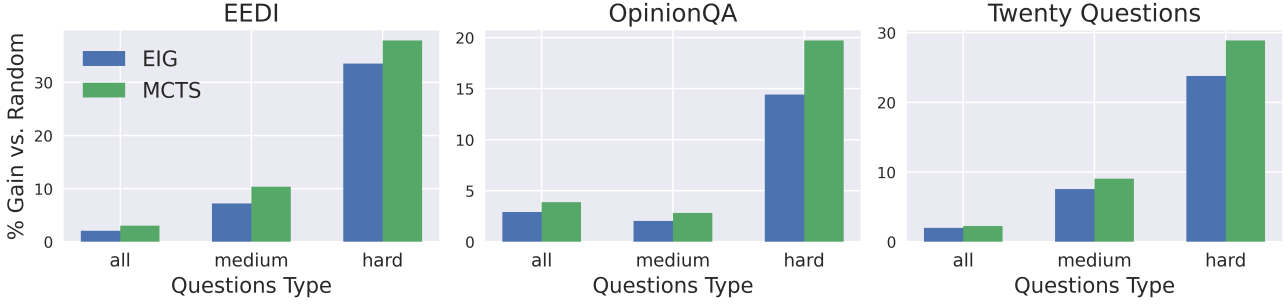


Figure 4. Relative accuracy gain from adaptive question selection (EIG and MCTS) over random selection for different subsets of target questions: all, medium difficulty (answer observed in  $< 50\%$  of the population), and hard (answer observed in  $< 30\%$ ). Adaptivity provides the greatest benefits when identifying rare latent traits, demonstrating when strategic question selection is most advantageous.

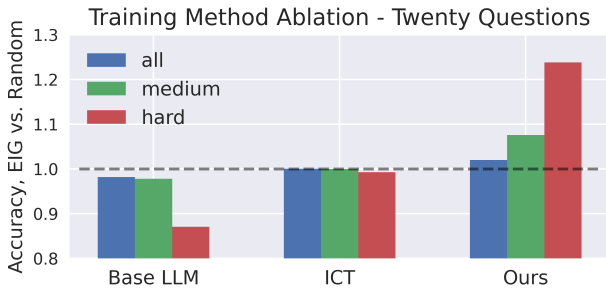


Figure 5. Comparison of performance gains from planning (EIG-based selection) using different models.

EIG or MCTS is more than 10x higher for hard questions than for all questions; for OpinionQA, it is 5x higher. We thus have strong evidence that our adaptive information elicitation strategy is most important when characterizing the latent features which are most atypical with respect to the population. If the latent entity exhibits atypical behavior (a student struggles with a concept that most find easy, or an opinion respondent holds a rare viewpoint), an adaptive method can target precisely those concepts that discriminate such cases. Conversely, random or fixed questionnaires fail to unearth those nuances within a limited query budget.

#### 4.4. Training Ablation

Our results in Figure 2 confirm the effectiveness of our end-to-end adaptive elicitation framework, while Figure 4 demonstrates the significant gains from planning-based question selection over random selection. Now, we turn to understanding the remaining component—meta-training—by evaluating how planning performs when applied to our model versus the ICT model and base LLM. We use the Twenty Questions dataset, and the same splits of all, medium, and hard questions as the previous experiment. For each setting and each of 3 underlying models, (Base, ICT, and ours), we record the accuracy on target questions after

selecting 3 questions with either random selection or the EIG strategy. To measure what is gained from planning, we record the ratio of target question accuracy with planning to that with random selection (a value above 1 indicates some accuracy gain from planning).

Results are shown in Figure 5. First, we see that planning performs poorly using the Base LLM, reducing accuracy almost 15% on hard questions compared to random question selection. The ICT model performance is largely unchanged by planning, across all 3 question types. On the other hand, our model’s performance is greatly improved when question selection is guided by planning, highlighting that our training procedure is essential to enable such strategic information gathering with LLMs.

#### 4.5. Other Ablations

We first ablate the number of candidate and target questions to choose from. Our experiments were run with the models being able to select from 20 questions in order to accurately predict 5 targets. In Table 2 in Appendix E, we find that our method gains more accuracy as the question bank becomes larger. In Table 1, we find that performance stays roughly the same as the number of target questions changes. Finally, we study the effect of the base model for our meta-training procedure. We test GPT2, Llama-3.2-1B, and Llama-3.1-8B, and find in Table 3 that performance increases as the model is larger.

### 5. Conclusion

In this work, we propose an adaptive elicitation framework, based on the missing-data view, that actively reduces uncertainty on the latent entity by simulating counterfactual responses. There is a rich body of literature on topics related to latent entity modeling and its applications, of which we are only able to give a limited overview here. See Appendix A for a thorough discussion of related research.

## Acknowledgements

We thank ONR Grant N00014-23-1-2436 for its generous support. This work is supported by the funds provided by the National Science Foundation and by DoD OUSD (R&E) under Cooperative Agreement PHY-2229929 (The NSF AI Institute for Artificial and Natural Intelligence).

## Impact Statement

The broader impact of this work includes significant potential for enhancing human-centered systems and services that require adaptive decision-making based on incomplete information. Our approach enables more targeted and personalized interactions, which could improve educational tools, assistive technologies, and personalized healthcare assessments. However, the ethical implications of such a framework must be carefully considered. Adaptive elicitation systems can shape user experiences, influence decision-making, and reinforce biases present in training data. We acknowledge the importance of ongoing scrutiny to ensure that such systems are deployed responsibly, particularly in high-stakes applications like healthcare and education.

## References

- Agrawal, S. and Goyal, N. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*, pp. 127–135. PMLR, 2013.
- Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., Das-Sarma, N., Drain, D., Fort, S., Ganguli, D., Henighan, T., Joseph, N., Kadavath, S., Kernion, J., Conerly, T., El-Showk, S., Elhage, N., Hatfield-Dodds, Z., Hernandez, D., Hume, T., Johnston, S., Kravec, S., Lovitt, L., Nanda, N., Olsson, C., Amodei, D., Brown, T., Clark, J., McCandlish, S., Olah, C., Mann, B., and Kaplan, J. Training a helpful and harmless assistant with reinforcement learning from human feedback, 2022. URL <https://arxiv.org/abs/2204.05862>.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022, March 2003. ISSN 1532-4435.
- Boutillier, C., Zemel, R. S., and Marlin, B. Active collaborative filtering. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence (UAI 2003)*, Toronto, ON, Canada, 2003.
- Brier, G. W. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. Language models are few-shot learners, 2020. URL <https://arxiv.org/abs/2005.14165>.
- Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S., and Colton, S. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1):1–43, 2012. doi: 10.1109/TCIAIG.2012.2186810.
- Castricato, L., Lile, N., Rafailov, R., Franken, J.-P., and Finn, C. Persona: A reproducible testbed for pluralistic alignment. *arXiv preprint*, arXiv:2407.17387, 2024. URL <https://arxiv.org/abs/2407.17387>.
- Chaloner, K. and Verdinelli, I. Bayesian Experimental Design: A Review. *Statistical Science*, 10(3):273 – 304, 1995. doi: 10.1214/ss/1177009939. URL <https://doi.org/10.1214/ss/1177009939>.
- Chapelle, O. and Li, L. An empirical evaluation of thompson sampling. In *Advances in Neural Information Processing Systems 24 (NeurIPS 2011)*, pp. 2249–2257. NeurIPS, 2011.
- Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., Laskin, M., Abbeel, P., Srinivas, A., and Mordatch, I. Decision transformer: Reinforcement learning via sequence modeling. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 15084–15097. Curran Associates, Inc., 2021. URL [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/7f48f642a0ddb10272b5c31057f0663-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/7f48f642a0ddb10272b5c31057f0663-Paper.pdf).
- Chen, Y., Zhong, R., Zha, S., Karypis, G., and He, H. Meta-learning via language model in-context tuning, 2022. URL <https://arxiv.org/abs/2110.07814>.
- Corbett, A. T. and Anderson, J. R. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4):253–278, 1994.
- Dawid, A. P. Statistical theory: The prequential approach. *Journal of the Royal Statistical Society, Series A*, 147: 278–292, 1984.
- de la Torre, J. The generalized dina model framework. *Psychometrika*, 76(2):179–199, 2011. doi: 10.1007/s11336-011-9207-7.

- DeepSeek-AI, Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., Zhang, X., Yu, X., Wu, Y., Wu, Z. F., Gou, Z., Shao, Z., Li, Z., Gao, Z., Liu, A., Xue, B., Wang, B., Wu, B., Feng, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., Dai, D., Chen, D., Ji, D., Li, E., Lin, F., Dai, F., Luo, F., Hao, G., Chen, G., Li, G., Zhang, H., Bao, H., Xu, H., Wang, H., Ding, H., Xin, H., Gao, H., Qu, H., Li, H., Guo, J., Li, J., Wang, J., Chen, J., Yuan, J., Qiu, J., Li, J., Cai, J. L., Ni, J., Liang, J., Chen, J., Dong, K., Hu, K., Gao, K., Guan, K., Huang, K., Yu, K., Wang, L., Zhang, L., Zhao, L., Wang, L., Zhang, L., Xu, L., Xia, L., Zhang, M., Zhang, M., Tang, M., Li, M., Wang, M., Li, M., Tian, N., Huang, P., Zhang, P., Wang, Q., Chen, Q., Du, Q., Ge, R., Zhang, R., Pan, R., Wang, R., Chen, R. J., Jin, R. L., Chen, R., Lu, S., Zhou, S., Chen, S., Ye, S., Wang, S., Yu, S., Zhou, S., Pan, S., Li, S. S., Zhou, S., Wu, S., Ye, S., Yun, T., Pei, T., Sun, T., Wang, T., Zeng, W., Zhao, W., Liu, W., Liang, W., Gao, W., Yu, W., Zhang, W., Xiao, W. L., An, W., Liu, X., Wang, X., Chen, X., Nie, X., Cheng, X., Liu, X., Xie, X., Liu, X., Yang, X., Li, X., Su, X., Lin, X., Li, X. Q., Jin, X., Shen, X., Chen, X., Sun, X., Wang, X., Song, X., Zhou, X., Wang, X., Shan, X., Li, Y. K., Wang, Y. Q., Wei, Y. X., Zhang, Y., Xu, Y., Li, Y., Zhao, Y., Sun, Y., Wang, Y., Yu, Y., Zhang, Y., Shi, Y., Xiong, Y., He, Y., Piao, Y., Wang, Y., Tan, Y., Ma, Y., Liu, Y., Guo, Y., Ou, Y., Wang, Y., Gong, Y., Zou, Y., He, Y., Xiong, Y., Luo, Y., You, Y., Liu, Y., Zhou, Y., Zhu, Y. X., Xu, Y., Huang, Y., Li, Y., Zheng, Y., Zhu, Y., Ma, Y., Tang, Y., Zha, Y., Yan, Y., Ren, Z. Z., Ren, Z., Sha, Z., Fu, Z., Xu, Z., Xie, Z., Zhang, Z., Hao, Z., Ma, Z., Yan, Z., Wu, Z., Gu, Z., Zhu, Z., Liu, Z., Li, Z., Xie, Z., Song, Z., Pan, Z., Huang, Z., Xu, Z., Zhang, Z., and Zhang, Z. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. URL <https://arxiv.org/abs/2501.12948>.
- Diao, S., Wang, P., Lin, Y., and Zhang, T. Active prompting with chain-of-thought for large language models, 2023.
- Du, Y., Yang, M., Florence, P., Xia, F., Wahid, A., Ichter, B., Sermanet, P., Yu, T., Abbeel, P., Tenenbaum, J. B., Kaelbling, L., Zeng, A., and Tompson, J. Video language planning. In *Proceedings of the International Conference on Learning Representations (ICLR)*. ICLR, 2024. Google Deepmind, Massachusetts Institute of Technology, UC Berkeley.
- Duan, J., Cheng, H., Wang, S., Zavalny, A., Wang, C., Xu, R., Kailkhura, B., and Xu, K. Shifting attention to relevance: Towards the predictive uncertainty quantification of free-form large language models, 2024.
- Embretson, S. E. and Reise, S. P. *Item Response Theory*. Psychology Press, New York, NY, 2013.
- Fong, E., Holmes, C., and Walker, S. G. Martingale posterior distributions. *Journal of the Royal Statistical Society, Series B*, 2023.
- Frazier, P. I. A tutorial on bayesian optimization. *arXiv preprint*, arXiv:1807.02811, 2018.
- Frazier, P. I., Powell, W. B., and Dayanik, S. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439, 2008. doi: 10.1137/070693424.
- Ghavamzadeh, M., Mannor, S., Pineau, J., and Tamar, A. Bayesian reinforcement learning: A survey. *Foundations and Trends in Machine Learning*, 8(5-6):359–483, 2015.
- Glushkova, T., Zerva, C., Rei, R., and Martins, A. F. T. Uncertainty-aware machine translation evaluation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.findings-emnlp.330. URL <http://dx.doi.org/10.18653/v1/2021.findings-emnlp.330>.
- Gonzalez, J., Osborne, M., and Lawrence, N. Glasses: Relieving the myopia of bayesian optimisation. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 2016.
- Grover, A., Markov, T., Attia, P., Jin, N., Perkins, N., Cheong, B., Chen, M., Yang, Z., Harris, S., Chueh, W., and Ermon, S. Best arm identification in multi-armed bandits with delayed feedback. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, pp. 833–842. PMLR, 2018.
- Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. On calibration of modern neural networks, 2017.
- Handa, K., Gal, Y., Pavlick, E., Goodman, N., Andreas, J., Tamkin, A., and Li, B. Z. Bayesian preference elicitation with language models, 2024. URL <https://arxiv.org/abs/2403.05534>.
- Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., and Baker, C. I. Things: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, 14(10):1–24, 10 2019. doi: 10.1371/journal.pone.0223792. URL <https://doi.org/10.1371/journal.pone.0223792>.
- Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., and Steinhardt, J. Measuring massive multitask language understanding. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

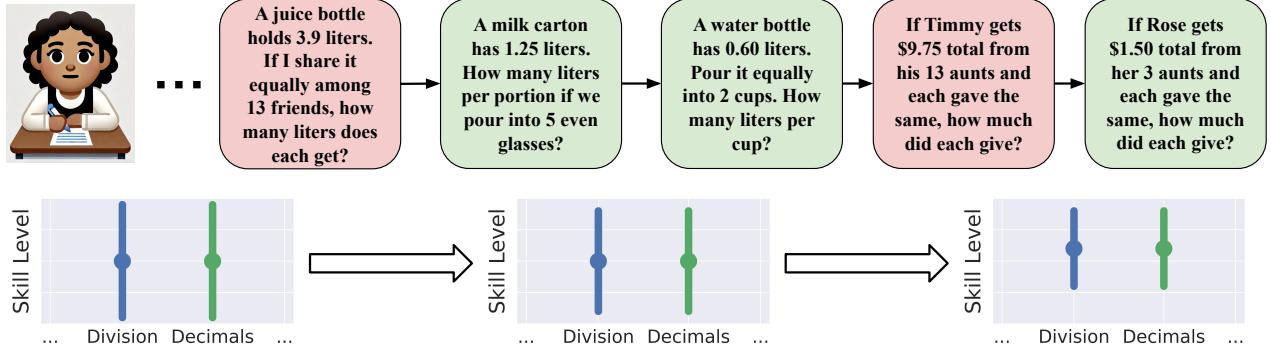
- Hill, B. M. Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 63(322):677–691, 1968.
- Hou, B., Liu, Y., Qian, K., Andreas, J., Chang, S., and Zhang, Y. Decomposing uncertainty for large language models through input clarification ensembling, 2024.
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. Lora: Low-rank adaptation of large language models, 2021. URL <https://arxiv.org/abs/2106.09685>.
- Hu, Z., Liu, C., Feng, X., Zhao, Y., Ng, S.-K., Luu, A. T., He, J., Koh, P. W., and Hooi, B. Uncertainty of thoughts: Uncertainty-aware planning enhances information seeking in large language models, 2024.
- Jang, J., Kim, S., Lin, B. Y., Wang, Y., Hessel, J., Zettlemoyer, L., Hajishirzi, H., Choi, Y., and Ammanabrolu, P. Personalized soups: Personalized large language model alignment via post-hoc parameter merging. *arXiv preprint*, arXiv:2310.11564, 2023. URL <https://arxiv.org/abs/2310.11564>.
- Janner, M., Li, Q., and Levine, S. Offline reinforcement learning as one big sequence modeling problem. In *Advances in neural information processing systems*, volume 34, pp. 1273–1286, 2021.
- Jelinek, F., Mercer, R. L., Bahl, L. R., and Baker, J. M. Perplexity—a measure of the difficulty of speech recognition tasks. *Journal of the Acoustical Society of America*, 62, 1977. URL <https://api.semanticscholar.org/CorpusID:121680873>.
- Jiang, S., Jiang, D., Balandat, M., Karrer, B., Gardner, J., and Garnett, R. Efficient nonmyopic bayesian optimization via one-shot multi-step trees. In *Advances in Neural Information Processing Systems 20*, 2020.
- Jones, D. R., Schonlau, M., and Welch, W. J. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998. ISSN 1573-2916. doi: 10.1023/A:1008306431147. URL <http://dx.doi.org/10.1023/A:1008306431147>.
- Jun, K.-S., Jamieson, K., Nowak, R., and Zhu, X. Top arm identification in multi-armed bandits with batch arm pulls. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, pp. 139–148. PMLR, 2016.
- Kadavath, S., Conerly, T., Askell, A., Henighan, T., Drain, D., Perez, E., Schiefer, N., Hatfield-Dodds, Z., DasSarma, N., Tran-Johnson, E., Johnston, S., El-Showk, S., Jones, A., Elhage, N., Hume, T., Chen, A., Bai, Y., Bowman,
- S., Fort, S., Ganguli, D., Hernandez, D., Jacobson, J., Kernion, J., Kravec, S., Lovitt, L., Ndousse, K., Olsson, C., Ringer, S., Amodei, D., Brown, T., Clark, J., Joseph, N., Mann, B., McCandlish, S., Olah, C., and Kaplan, J. Language models (mostly) know what they know, 2022.
- Kirk, H. R., Whitefield, A., Röttger, P., Bean, A., Margatina, K., Ciro, J., Mosquera, R., Bartolo, M., Williams, A., He, H., Vidgen, B., and Hale, S. A. The prism alignment project: What participatory, representative and individualised human feedback reveals about the subjective and multicultural alignment of large language models. *arXiv preprint*, arXiv:2404.16019, 2024. URL <https://arxiv.org/abs/2404.16019>.
- Kuhn, L., Gal, Y., and Farquhar, S. Semantic uncertainty: Linguistic invariances for uncertainty estimation in natural language generation, 2023.
- Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge, 2019.
- Lee, J., Xie, A., Pacchiano, A., Chandak, Y., Finn, C., Nachum, O., and Brunskill, E. In-context decision-making from supervised pretraining. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023. URL <https://openreview.net/forum?id=WIZyLD6j6E>.
- Lee, K.-H., Nachum, O., Yang, M., Lee, L., Freeman, D., Xu, W., Guadarrama, S., Fischer, I., Jang, E., Michalewski, H., and Mordatch, I. Multi-game decision transformers. In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS)*. NeurIPS, 2022.
- Li, J., Mehrabi, N., Peris, C., Goyal, P., Chang, K.-W., Galstyan, A., Zemel, R., and Gupta, R. On the steerability of large language models toward data-driven personas, 2024a. URL <https://arxiv.org/abs/2311.04978>.
- Li, L., Chu, W., Langford, J., and Schapire, R. E. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW 2010)*, pp. 661–670. ACM, 2010.
- Li, X., Lipton, Z. C., and Leqi, L. Personalized language modeling from personalized human feedback. *arXiv preprint*, arXiv:2402.05133, 2024b. URL <https://arxiv.org/abs/2402.05133>.
- Lin, L., Bai, Y., and Mei, S. Transformers as decision makers: Provable in-context reinforcement learning via supervised pretraining. In *The Twelfth International Conference on Learning Representations*, 2024a. URL <https://openreview.net/forum?id=yN4Wv17ss3>.



- Lin, Z., Liu, J. Z., and Shang, J. Towards collaborative neural-symbolic graph semantic parsing via uncertainty. In Muresan, S., Nakov, P., and Villavicencio, A. (eds.), *Findings of the Association for Computational Linguistics: ACL 2022*, pp. 4160–4173, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-acl.328. URL <https://aclanthology.org/2022.findings-acl.328>.
- Lin, Z., Trivedi, S., and Sun, J. Generating with confidence: Uncertainty quantification for black-box large language models, 2024b.
- Lindley, D. V. *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Cambridge University Press, 1965.
- Liu, J., Shen, D., Zhang, Y., Dolan, B., Carin, L., and Chen, W. What makes good in-context examples for gpt-3?, 2021. URL <https://arxiv.org/abs/2101.06804>.
- Liu, Q., Zhuang, Y., Bi, H., Huang, Z., Huang, W., Li, J., Yu, J., Liu, Z., Hu, Z., Hong, Y., Pardos, Z. A., Ma, H., Zhu, M., Wang, S., and Chen, E. Survey of computerized adaptive testing: A machine learning perspective. *arXiv preprint arXiv:2404.00712*, April 2024. URL <https://arxiv.org/abs/2404.00712>.
- Loshchilov, I. and Hutter, F. Decoupled weight decay regularization, 2019. URL <https://arxiv.org/abs/1711.05101>.
- Malinin, A. and Gales, M. Uncertainty estimation in autoregressive structured prediction, 2021.
- Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14:265–294, 1978.
- Osband, I., Blundell, C., Pritzel, A., and Roy, B. V. Deep exploration via bootstrapped dqn. In *Advances in Neural Information Processing Systems*, volume 29, 2016.
- Osband, I., Aslanides, J., and Cassirer, A. Randomized prior functions for deep reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 31, 2018.
- Osband, I., Wen, Z., Asghari, S. M., Dwaracherla, V., Lu, X., Ibrahim, M., Lawson, D., Hao, B., O’Donoghue, B., and Van Roy, B. The neural testbed: Evaluating joint predictions. In *Advances in Neural Information Processing Systems*, volume 35, pp. 12554–12565, 2022.
- Osband, I., Asghari, S. M., Roy, B. V., McAleese, N., Aslanides, J., and Irving, G. Fine-tuning language models via epistemic neural networks, 2023a.
- Osband, I., Wen, Z., Asghari, S. M., Dwaracherla, V., Ibrahim, M., Lu, X., and Van Roy, B. Approximate thompson sampling via epistemic neural networks. In *Uncertainty in Artificial Intelligence*, pp. 1586–1595. PMLR, 2023b.
- Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., and Bernstein, M. S. Generative agent simulations of 1,000 people. 2024. Preprint.
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., and Sohl-Dickstein, J. Deep knowledge tracing. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL [https://proceedings.neurips.cc/paper\\_files/paper/2015/file/bac9162b47c56fc8a4d2a519803d51b3-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/bac9162b47c56fc8a4d2a519803d51b3-Paper.pdf).
- Poddar, S., Wan, Y., Ivison, H., Gupta, A., and Jaques, N. Personalizing reinforcement learning from human feedback with variational preference learning. 2024.
- Qin, C., Wen, Z., Lu, X., and Van Roy, B. An analysis of ensemble sampling. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 21602–21614. Curran Associates, Inc., 2022.
- Rafferty, A. N., Brunskill, E., Griffiths, T. L., and Shafto, P. Faster teaching by pomdp planning. In *Artificial Intelligence in Education*, pp. 280–287. Springer, 2011.
- Riquelme, C., Tucker, G., and Snoek, J. Deep bayesian bandits showdown: An empirical comparison of bayesian deep networks for thompson sampling. In *International Conference on Learning Representations*, 2018.
- Rubin, D. B. Inference and missing data. *Biometrika*, 63(3):581–592, 1976. ISSN 00063444, 14643510. URL <http://www.jstor.org/stable/2335739>.
- Russo, D. Simple bayesian algorithms for best-arm identification. *Operations Research*, 68(6):1625–1647, 2020.
- Russo, D., Van Roy, B., Kazerouni, A., Osband, I., and Wen, Z. A tutorial on thompson sampling. 2020.
- Ryan, E. G., Drovandi, C. C., McGree, J. M., and Pettitt, A. N. A review of modern computational algorithms for bayesian optimal design. *International Statistics Review*, 84(1):128–154, 2016.

- Salakhutdinov, R. and Mnih, A. Probabilistic matrix factorization. In *Proceedings of the 21st International Conference on Neural Information Processing Systems, NIPS'07*, pp. 1257–1264, Red Hook, NY, USA, 2007. Curran Associates Inc. ISBN 9781605603520.
- Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., and Hashimoto, T. Whose opinions do language models reflect? *arXiv preprint arXiv:2303.17548*, 2023.
- Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., and Wang, L. Prompting gpt-3 to be reliable, 2023.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, January 2016. doi: 10.1038/nature16961.
- Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M., Prabhat, M., and Adams, R. Scalable bayesian optimization using deep neural networks. In *International conference on machine learning*, pp. 2171–2180. PMLR, 2015.
- Torre, J. D. L. Dina model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1):115–130, 2009.
- Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., and Zhou, D. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*, 2022.
- Wang, Z., Lamb, A., Saveliev, E., Cameron, P., Zaykov, Y., Hernández-Lobato, J. M., Turner, R. E., Baraniuk, R. G., Barton, C., Jones, S. P., Woodhead, S., and Zhang, C. Diagnostic questions: The neurips 2020 education challenge. *arXiv preprint arXiv:2007.12061*, 2020.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24824–24837, 2022.
- Wen, Z., Osband, I., Qin, C., Lu, X., Ibrahimi, M., Dwaracherla, V., Asghari, M., and Roy, B. V. From predictions to decisions: The importance of joint predictive distributions. *arXiv preprint*, arXiv:2107.09224, 2021.
- Yadkori, Y. A., Kuzborskij, I., György, A., and Szepesvári, C. To believe or not to believe your llm, 2024.
- Yang, S., Nachum, O., Du, Y., Wei, J., Abbeel, P., and Schuurmans, D. Foundation models for decision making: Problems, methods, and opportunities. *arXiv preprint arXiv:2303.04129*, 2023.
- Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., and Narasimhan, K. Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601*, 2023.
- Ye, N., Yang, H., Siah, A., and Namkoong, H. Exchangeable sequence models can naturally quantify uncertainty over latent concepts. *arXiv preprint arXiv:2408.03307*, 2024. Available at <https://arxiv.org/abs/2408.03307>.
- Zhang, K. W., Cai, T. T., Namkoong, H., and Russo, D. Posterior sampling via autoregressive generation. *arXiv preprint arXiv:2405.19466*, 2024. Published on arXiv, 8 October 2024.
- Zollo, T. P., Siah, A. W. T., Ye, N., Li, A., and Namkoong, H. Personalllm: Tailoring llms to individual preferences, 2025. URL <https://arxiv.org/abs/2409.20296>.

## Static



## Adaptive

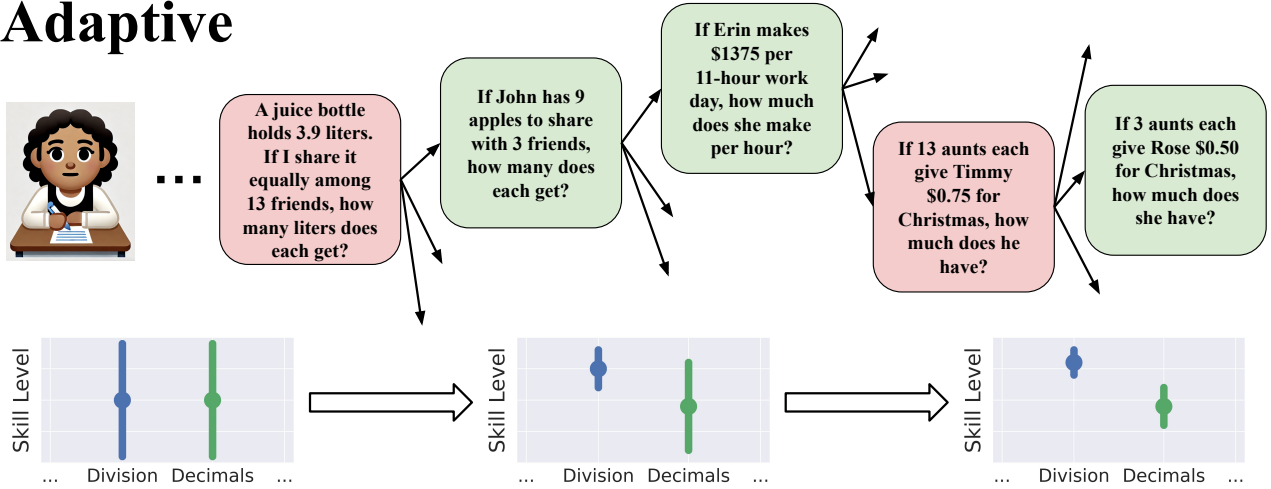


Figure 6. An example of how adaptive elicitation (bottom) can improve over static strategies (top) in new student assessment. Each question asked to the student is marked green (answered correctly) or red (answered incorrectly). Once a student answers incorrectly, the adaptive strategy is able to search over the *high-dimensional question space*, and present the student with a series of more granular examples that resolve uncertainty about specific abilities in the *high-dimensional space representing the student’s latent abilities*. In this case, active question selection reveals that the student is strong in division, but struggles with decimal points. A static assessment, on the other hand, fails to resolve this uncertainty.

## Reproducibility

Our code is available at <https://github.com/namkoong-lab/adaptive-elicitation>.

### A. Related Work

**Latent Uncertainty Modeling and Decision Making** Traditional works that attempt to model latent uncertainty to make robust decisions often pose explicit Bayesian models that directly specify a latent parameter. Classical multi-armed bandits such as Thompson Sampling-based methods (Russo, 2020; Grover et al., 2018; Chapelle & Li, 2011; Agrawal & Goyal, 2013; Lattimore & Szepesvári, 2019; Jun et al., 2016; Li et al., 2010) specify a Bayesian model as a prior, which can be used to draw explicit posterior samples. Examples of Bayesian models include Gaussian and Bernoulli distributions, as well as Bayesian linear or logistic regression (Russo et al., 2020). Bayesian Optimal Experimental Design (BOED) methods (Ghavamzadeh et al., 2015; Chaloner & Verdinelli, 1995; Ryan et al., 2016) follow a similar paradigm by quantifying information gain using explicit Bayesian models to make sequential decisions.

Other lines of work include more sophisticated probabilistic modeling. Active collaborative filtering methods (Boutillier et al., 2003) specify probabilistic models of user preference data. Bayesian Optimization techniques (Frazier et al., 2008; Frazier, 2018; Gonzalez et al., 2016; Jiang et al., 2020; Jones et al., 1998) apply acquisition functions such as Expected Improvement (Jones et al., 1998) and Knowledge Gradient (Frazier et al., 2008) on top of Gaussian Processes. While these traditional methods are statistically principled, the need to specify explicit models means they often struggle to model very high-dimensional spaces such as the space of natural language. For example, (Frazier, 2018) mentions that Bayesian Optimization methods are best suited to dimensionality  $\leq 20$ , while natural language embeddings are often thousands of dimensions.

To overcome these limitations, recent works have focused on representing uncertainty over high dimensions augmented by the representation power of neural networks. One line of works explicitly uses neural network representations to represent the underlying uncertainty, where decision making algorithms are applied on top of Bayesian Neural Networks (BNNs) or the last layer representation (Snoek et al., 2015; Riquelme et al., 2018; Osband et al., 2016; 2018; Piech et al., 2015). Another strand of works uses ensembles to represent the underlying uncertainty (Qin et al., 2022), or more efficient variants such as Epistemic Neural Networks (Wen et al., 2021; Osband et al., 2022; 2023b). We offer a different approach by directly modeling the uncertainty surrounding future predictions using language models. Methods that explicitly represent uncertainty in terms of ensembles or through specified parameters still struggle to operate in the discrete, high-dimensional space of natural language, whereas our perspective is able to directly represent uncertainty in natural language predictions. Additionally, our meta-learning method does not require new architectures and can be applied on top of powerful pre-trained language models, allowing the use of internet-scale language understanding in comprehending uncertainty.

**Computerized Adaptive Testing** Our work is closely related to Computerized Adaptive Testing (CAT) methods, a form of educational testing that adapts to a student’s ability level. Classical CAT methods attempt to capture a student’s latent ability level through simple parametrized models. Item Response Theory, also known as Latent Response Theory, includes a family of simple mathematical models such as a one parameter logistic regression or Item Response Function to model a student’s latent ability (Liu et al., 2024; Embretson & Reise, 2013). The Diagnostic Classification Model (DCM) is designed to measure proficiency across a wide array of specific knowledge concepts. A prominent example includes the DINA method (Torre, 2009; de la Torre, 2011) which uses a probabilistic binary matrix model to represent these concepts. Knowledge Tracing techniques, which train machine learning methods to model the latent knowledge of students as they interact with coursework, traditionally use Hidden Markov Models (HMMs) (Corbett & Anderson, 1994) or Partially Observable Markov Decision Processes (POMDPs) (Rafferty et al., 2011) to model the latent state. More modern treatments use deep learning models such as a Recurrent Neural Network (Piech et al., 2015) to represent latent knowledge.

**Reinforcement Learning with Sequence Models** A number of works propose to train or use powerful pre-trained models in order to solve complex reinforcement learning (RL) tasks, focusing on how these models can make decisions using vast amounts of offline data (Janner et al., 2021; Yang et al., 2023; Chen et al., 2021; Du et al., 2024; Lee et al., 2022). Another line of works show that using meta-learned sequence models to predict the next action can approximate standard bandit algorithms (Lin et al., 2024a; Lee et al., 2023; Zhang et al., 2024). We extend these ideas to natural language while focusing on how our meta-learned model can quantify uncertainty to make a decision.

**Uncertainty Quantification over Natural Language.** There has been a recent class of works focusing on developing uncertainty measures to augment the reliability of model responses. (Kuhn et al., 2023; Lin et al., 2024b; Malinin & Gales, 2021; Duan et al., 2024) focus on predictive entropy measures with off-the-shelf language models, while other approaches focus on self-consistency in the generation space (Lin et al., 2022; Si et al., 2023; Kadavath et al., 2022; Diao et al., 2023). Another class of works focuses instead on detecting *epistemic* uncertainty from *aleatoric* uncertainty in model outputs (Yadkori et al., 2024; Osband et al., 2023a; Hou et al., 2024; Glushkova et al., 2021). Our meta-learning uncertainty quantification framework is complementary to these works, as these measures are designed to be applied on top of pre-trained foundation models.

**Planning and Information Gathering with LLMs** Our work is related to Uncertainty of Thoughts (UoT) (Hu et al., 2024) and OPEN (Handa et al., 2024). While these methods build elicitation procedures on top of off-the-shelf language models, we use a meta-learning procedure in order to accurately quantify uncertainty over new environments. Other works introduce methods to enhance general reasoning or planning capabilities by using natural language reasoning steps (Wei et al., 2022; Wang et al., 2022; Yao et al., 2023).



**Personalization with Language Models** With the recent success of Large Language Models (LLMs), a natural question is whether these models can be tailored and personalize to various users. There has been a nascent series of works that propose benchmarks and methods for personalized language modeling. (Zollo et al., 2025; Kirk et al., 2024; Castricato et al., 2024) propose new testbeds and evaluation criteria that target various dimensions of personalization through synthetic and real data. (Jang et al., 2023) proposes to merge model parameters to personalize models, while (Li et al., 2024b; Poddar et al., 2024) propose new personalized fine-tuning and Reinforcement Learning from Human Feedback (RLHF) techniques. In the context of opinion polling, (Li et al., 2024a) steers model outputs to different personas by using embeddings from collaborative filtering, while (Park et al., 2024) demonstrates that language models can successfully be adapted to individual responses.

## B. Theoretical Validity

We now detail two propositions that provide insight into our meta-learning framework. The first proposition lower bounds performance of our model in terms of the performance in the ideal simulation as well as the difference between the true and simulated distribution. The second proposition quantifies the performance gap between a greedy and full combinatorial method, showing the losses one might expect to incur.

First, let  $\mathcal{X}_T = (x_1, \dots, x_T)$ . Define  $\mathcal{X}^*$  to be the optimal set of questions  $X$  that maximizes the log likelihood of the object of interest  $Z$  under the meta-learned distribution  $p_\theta$ ,

$$\mathcal{X}^* := \operatorname{argmax}_{\mathcal{X}_T} \mathbb{E}_{p_\theta} [\log p_\theta(Z \mid \mathcal{X}_T)].$$

This is the set of questions we would ask if we could perfectly optimize in our simulated environment. Next, let  $q$  be the true distribution and  $p_\theta$  our meta-learned model. Using these definitions, we state our first proposition

**Proposition B.1.** *For any  $p_\theta$ ,*

$$\mathbb{E}_q [\log p_\theta(Z \mid \mathcal{X}^*)] \geq \mathbb{E}_{p_\theta} [\log p_\theta(Z \mid \mathcal{X}^*)] - \sqrt{\mathbb{E}_{p_\theta} [\log^2 p_\theta(Z \mid \mathcal{X}^*)] \chi^2(q(Z) \parallel p_\theta(Z \mid \mathcal{X}^*))}. \quad (9)$$

$\mathbb{E}_q [\log p_\theta(Z \mid \mathcal{X}^*)]$  represents the cross entropy of  $Z$  between  $p_\theta$  and the ground truth distribution  $q$ . The lower bound first involves  $\mathbb{E}_{p_\theta} [\log p_\theta(Z \mid \mathcal{X}^*)]$ , which is the likelihood of the data under our simulated distribution. The second term involves both the likelihood under the simulated distribution  $\mathbb{E}_{p_\theta} [\log^2 p_\theta(Z \mid \mathcal{X}^*)]$  as well as the distance between  $q$  and  $p$  through  $\chi^2(q(Z) \parallel p_\theta(Z \mid \mathcal{X}^*))$ . This bound tells us that if  $p_\theta$  has high likelihood under the simulator and has little distance to  $q$ , then we are guaranteed to achieve good test-time performance. However, since the difference  $\chi^2$  is scaled by the likelihood, simply having high likelihood in the simulated distribution is not enough. In fact, if  $\chi^2(q \parallel p)$  is large, then having high likelihood in the simulation can exacerbate this error in the second term. We prove this in Appendix B.2.

**Bound for Greedy Policy** Next, we bound the performance gap between the questions collected from the optimal policy  $\mathcal{X}^*$  with those collected from the greedy policy

$$x_i^{\text{greedy}} = \operatorname{argmax}_{x_i} \mathbb{E}_{p_\theta} [\log p_\theta(Z \mid (x_1, \dots, x_{i-1}) \cup x_i)],$$

where  $\mathcal{X}_{\text{greedy}} = (x_1^{\text{greedy}}, \dots, x_T^{\text{greedy}})$ . First we define the concept of submodularity:

**Definition B.2** (Submodularity).  $f : 2^\Omega \rightarrow \mathbb{R}$  is *submodular* if  $\forall X \subseteq Y \subseteq \Omega$  and  $\forall z \notin Y$  we have

$$f(X \cup \{z\}) - f(X) \geq f(Y \cup \{z\}) - f(Y)$$

Then we state our proposition:

**Proposition B.3.** *Under the assumption that the entropy over  $Z$  produced by the meta-learned model  $p_\theta$  is submodular,*

$$\mathbb{E}_{p_\theta} [\log p_\theta(Z \mid \mathcal{X}^*)] - \mathbb{E}_{p_\theta} [\log p_\theta(Z \mid \mathcal{X}_{\text{greedy}})] \leq \frac{1}{e} \text{EIG}(Z; \mathcal{X}^*).$$

We prove this statement in Appendix B.3. This proposition states that the difference in achieving the optimal log likelihood under the simulated environment and using the greedy strategy is bounded by  $\frac{1}{e} \text{EIG}(Z; \mathcal{X}^*)$ . We can then use this bound

and substitute in Proposition B.1 to quantify the performance lower bound for the greedy policy. In this proposition, we have to assume submodularity of our meta-learned model because it is not guaranteed in practice due to training instabilities or errors. Empirically we find that the entropy of our meta-learned model behaves submodularly, as evidenced by the perplexity graphs in Figure 2.

### B.1. Summary

Our framework provides a data-driven, natural-language-based alternative to parametric modeling of latent entities. It proceeds by: (1) *Meta-training* a language model on diverse question–answer sequences, (2) Interpreting the model’s predictive distribution over future answers as an *uncertainty measure* about new entities, and (3) Iteratively *selecting questions* to optimally reduce that uncertainty. We show theoretically that our procedure gives strong performance, even under a simple and efficient greedy planning strategy. Next, we explore the empirical performance of our framework in a series of adaptive information gathering scenarios.

In this section, we show the theoretical validity of using a greedy procedure to select actions with the highest expected information gain. We wish to quantify and reduce our uncertainty about some object  $Z$  by choosing the optimal questions  $X$  to query the latent entity  $U$ . First we define the expected information gain of asking a set of questions  $\mathcal{X}_t := (x_1, \dots, x_t) \subseteq \mathcal{X}$ ,

$$\text{EIG}(Z; \mathcal{X}_t) = H(Z) - \mathbb{E}_t \left[ H \left( Z \mid \bigcup_{s=1}^t (x_s, Y_s) \right) \right].$$

Note that each  $Y_s$  in the history is a random variable, and we use our meta-learned model  $p_\theta$  to simulate possible answers  $Y_s \sim p_\theta(\cdot | \mathcal{H}_{s-1}, X_s = x_s)$ . Similarly,  $Z \sim p_\theta(\cdot)$  as well. Ultimately, our goal is to choose a set of designs  $\mathcal{X}_t = x_{1:t}$  that yields the largest amount of information gain possible. First to set notation, define  $q$  to be the ground truth underlying question and answer distribution. Let  $p$  be the distribution induced by the meta-learned model, and  $p(Z | \mathcal{X}_t)$  be the conditional distribution over  $Z$  after marginalizing out the feedback  $Y_{1:t}$  corresponding to the questions  $\mathcal{X}_t$ , where  $Y_{1:t}$  comes from the ground truth distribution such that  $Y_{1:t} \sim q(\cdot)$ .

First, define  $\mathcal{X}^*$  to be the optimal set of questions  $X$  that maximizes the log likelihood of the object of interest  $Z$  under the meta-learned distribution  $P$ ,

$$\mathcal{X}^* := \underset{\mathcal{X}_t}{\operatorname{argmax}} \mathbb{E}_p[\log p(Z | \mathcal{X}_t)].$$

### B.2. Proof of Proposition B.1

We restate our proposition for clarity.

$$\mathbb{E}_q[\log p(Z | \mathcal{X}^*)] \geq \mathbb{E}_p[\log p(Z | \mathcal{X}^*)] - \sqrt{\mathbb{E}_p[\log^2 p(Z | \mathcal{X}^*)] \chi^2(q(Z) \parallel p(Z | \mathcal{X}^*))}.$$

*Proof.* We can decompose the difference between the cross entropy term  $\mathbb{E}_q[\log p(Z | \mathcal{X}^*)]$  and entropy term  $\mathbb{E}_p[\log p(Z | \mathcal{X}^*)]$  as

$$\begin{aligned} & -\mathbb{E}_q[\log p(Z | \mathcal{X}^*)] + \mathbb{E}_p[\log p(Z | \mathcal{X}^*)] \\ &= \int (p(z | \mathcal{X}^*) - q(z)) \log p(z | \mathcal{X}^*) dz \\ &= \int \left(1 - \frac{q(z)}{p(z | \mathcal{X}^*)}\right) \log p(z | \mathcal{X}^*) p(z | \mathcal{X}^*) dz \\ &\leq \sqrt{\left(\int \left(1 - \frac{q(z)}{p(z | \mathcal{X}^*)}\right)^2 p(z | \mathcal{X}^*) dz\right) \left(\int \log^2(p(z | \mathcal{X}^*)) p(z | \mathcal{X}^*) dz\right)} \\ &= \sqrt{\mathbb{E}_p[\log^2 p(Z | \mathcal{X}^*)] \chi^2(q(Z) \parallel p(Z | \mathcal{X}^*))}. \end{aligned}$$

where the second to last line follows from the Cauchy-Schwartz inequality and the last line follows from the definition of the  $\chi^2$  divergence. Then by flipping the sign, we obtain the stated inequality.  $\square$

### B.3. Proof of Proposition B.3

We first define submodularity for clarity:

**Definition B.4** (Submodularity).  $f : 2^\Omega \rightarrow \mathbb{R}$  is *submodular* if  $\forall X \subseteq Y \subseteq \Omega$  and  $\forall z \notin Y$  we have

$$f(X \cup \{z\}) - f(X) \geq f(Y \cup \{z\}) - f(Y)$$

In order to show that the greedy procedure is able to perform close to the optimal solution, we rely on the following assumption, that the entropy calculated from our meta-learned model is submodular:

**Assumption B.5** (Submodularity of Entropy). Let  $\mathcal{H}_t \subseteq \mathcal{H}'_t$ . Then for any  $(X_{t+1}, Y_{t+1}) \notin \mathcal{H}'_t$ ,

$$H(Z | \mathcal{H}_t \cup (X_{t+1}, Y_{t+1})) - H(Z | \mathcal{H}_t) \geq H(Z | \mathcal{H}'_t \cup (X_{t+1}, Y_{t+1})) - H(Z | \mathcal{H}'_t).$$

We first show that the Expected Information Gain (EIG) is submodular. If the entropy is submodular, then the Information Gain is also submodular. Define the Information Gain as

$$f(\mathcal{H}_t) = \text{IG}(Z; \mathcal{H}_t) = H(Z) - H(Z | \mathcal{H}_t).$$

Then, for any history set  $\mathcal{H}_t$  and any additional observation  $(X_{t+1}, Y_{t+1}) \notin \mathcal{H}_t$ , the marginal gain of adding  $(X_{t+1}, Y_{t+1})$  is given by

$$\begin{aligned} f(\mathcal{H}_t \cup \{(X_{t+1}, Y_{t+1})\}) - f(\mathcal{H}_t) &= \left[ H(Z) - H(Z | \mathcal{H}_t \cup \{(X_{t+1}, Y_{t+1})\}) \right] - \left[ H(Z) - H(Z | \mathcal{H}_t) \right] \\ &= H(Z | \mathcal{H}_t) - H(Z | \mathcal{H}_t \cup \{(X_{t+1}, Y_{t+1})\}). \end{aligned}$$

Now, consider two history sets  $\mathcal{H}_t \subseteq \mathcal{H}'_t$  and the same observation  $(X_{t+1}, Y_{t+1}) \notin \mathcal{H}'_t$ . By our submodularity assumption on the entropy, we have

$$H(Z | \mathcal{H}_t) - H(Z | \mathcal{H}_t \cup \{(X_{t+1}, Y_{t+1})\}) \geq H(Z | \mathcal{H}'_t) - H(Z | \mathcal{H}'_t \cup \{(X_{t+1}, Y_{t+1})\}).$$

In terms of the Information Gain function, this inequality becomes

$$f(\mathcal{H}_t \cup \{(X_{t+1}, Y_{t+1})\}) - f(\mathcal{H}_t) \geq f(\mathcal{H}'_t \cup \{(X_{t+1}, Y_{t+1})\}) - f(\mathcal{H}'_t).$$

Thus, by definition the Information Gain is submodular. Since this is true for all  $X_{t+1}, Y_{t+1}$ , then the Expected Information Gain (EIG) is also submodular. Then by (Nemhauser et al., 1978), we have that

$$\text{EIG}(Z; \mathcal{X}_{\text{greedy}}) \geq (1 - \frac{1}{e})\text{EIG}(Z; \mathcal{X}^*),$$

implying that

$$\text{EIG}(Z; \mathcal{X}_{\text{greedy}}) - \text{EIG}(Z; \mathcal{X}^*) \leq \frac{1}{e}\text{EIG}(Z; \mathcal{X}^*). \quad (10)$$

Finally, to prove the bound we can note that

$$\begin{aligned} \mathbb{E}_p[\log p(Z | \mathcal{X}^*)] - \mathbb{E}_p[\log p(Z | \mathcal{X}_{\text{greedy}})] &= \mathbb{E}_Y[\mathbb{E}_Z[\log p(Z | \mathcal{X}^*, \mathcal{Y}^*)]] - \mathbb{E}_Y[\mathbb{E}_Z[\log p(Z | \mathcal{X}_{\text{greedy}}, \mathcal{Y}_{\text{greedy}})]] \\ &= \mathbb{E}_Y[H(Z | \mathcal{X}^*, \mathcal{Y}^*)] - \mathbb{E}_Y[H(Z | \mathcal{X}_{\text{greedy}}, \mathcal{Y}_{\text{greedy}})] \\ &= \mathbb{E}_Y[H(Z | \mathcal{X}^*, \mathcal{Y}^*)] - H(Z) + H(Z) - \mathbb{E}_Y[H(Z | \mathcal{X}_{\text{greedy}}, \mathcal{Y}_{\text{greedy}})] \\ &= \text{EIG}(Z; \mathcal{X}_{\text{greedy}}) - \text{EIG}(Z; \mathcal{X}^*) \\ &\leq \frac{1}{e}\text{EIG}(Z; \mathcal{X}^*), \end{aligned}$$

where the last line follows from an application of Equation (10). This completes the proof.

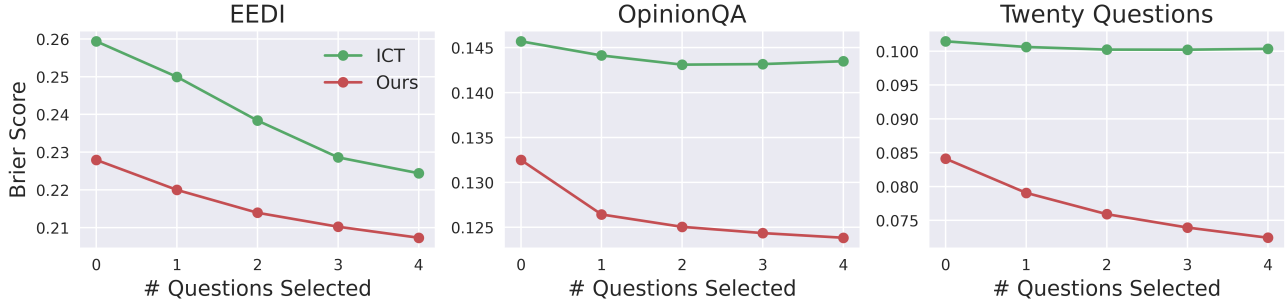


Figure 7. Brier score results in our overall setting across 3 datasets.

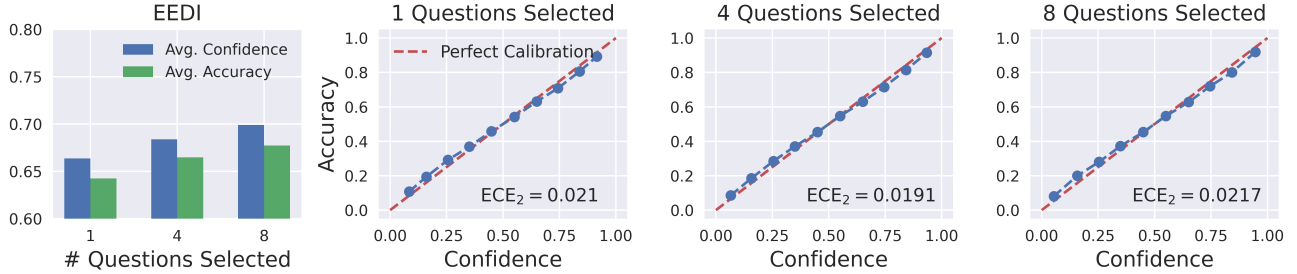


Figure 8. Calibration results with EEDI.

## C. Experiment Details

### C.1. Meta-Training Details

We initialize the AdamW (Loshchilov & Hutter, 2019) optimizer with learning rate of 0.0001 and  $\beta = (0.9, 0.95)$ , weight decay of 0.1, and we use a linear warmup for the learning rate after which we use a cosine scheduler. We train our model for 10,000 epochs with a batch size of 4 and block size of 1024, after which we take the checkpoint with the lowest validation loss.

## D. Experiment Results

Here we include additional experiment results. Figure 7 shows results for the overall experiments with the Brier Score metric. Figure 8 shows calibration results for EEDI, and Figure 9 shows calibration results for Twenty Questions.

## E. Ablations

Additional ablations are shown in Table 1 (number of target questions), Table 2 (number of candidate questions), Table 3 (base model).

Table 1. Ablating number of target questions on EEDI; accuracy conditioned on 4 Questions

Accuracy	1	5	10	20
Base	0.6042	0.6005	0.6066	0.5987
Ictx	0.6269	0.6278	0.6295	0.6255
Ours	<b>0.6759</b>	<b>0.6784</b>	<b>0.6871</b>	<b>0.6832</b>



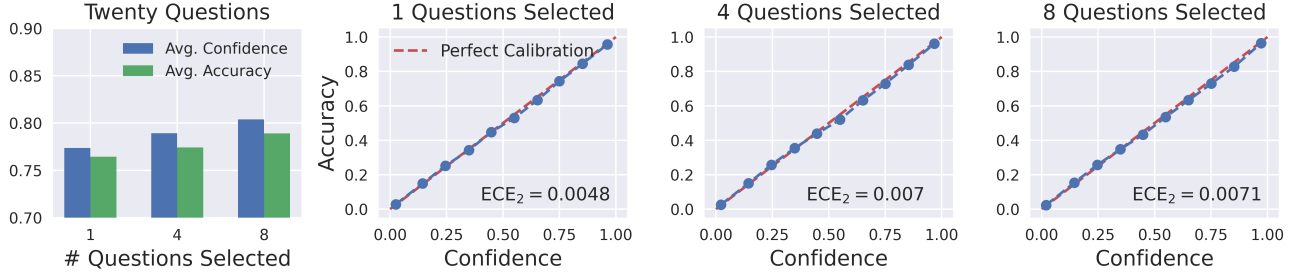


Figure 9. Calibration results with Twenty Questions.

Table 2. Ablating number of candidate questions on OpinionQA; accuracy conditioned on 4 Questions

Accuracy	10	15	20	25
Base	0.4030	0.4042	0.4089	0.4093
Ictx	0.4988	0.4993	0.5023	0.5009
Ours	<b>0.5933</b>	<b>0.5953</b>	<b>0.5987</b>	<b>0.6068</b>

Table 3. Ablating base model: Twenty Questions accuracy using our framework conditioned on 4 questions

	GPT2	Llama-3.2-1B	Llama-3.1-8B
Accuracy	0.5201	0.6131	0.7382