Bayesian Teaching Enables Probabilistic Reasoning in Large Language Models

Presenter: YunSeop Shin

May 20, 2026

Seoul national university, statistics, IDEA LAB

Motivation

- LLMs are increasingly used as agents that interact with users over multiple rounds (e.g., recommendation systems).
- To make good recommendations, the LLM must infer user preferences from observed choices and update its beliefs as new information arrives.
- The Bayesian inference framework defines the normative way to perform such belief updates.
- It starts by assigning uniform probability over user preferences, and as observations accumulate, it gradually concentrates probability mass on specific preferences.
- **Key question:** Can LLMs be made to behave as if they follow a Bayesian framework?

Note

- This paper does not quantify the uncertainty of LLM outputs.
- Here, probabilistic reasoning means that the LLM learns to update its beliefs in light of new observations, in a manner that mimics Bayesian inference.
- Actually, the LLM does not explicitly compute any posterior distribution.

Task: Flight Recommendation

- The paper mainly considers flight recommendation task.
- A simulated user interacts with an assistant for 5 rounds.
- In each round, $k = 3$ flight options are presented.
- The user selects a flight based on hidden preferences.
- The assistant observes the user's choice and must recommend flights that better match the user in subsequent rounds.
- The user's preferences are not directly given. They must be inferred from choices.

Notation

- $O = \{o_1, \ldots, o_k\}$: Set of k options ($k = 3$).
- o : An individual option.
- O_i: Set of options in round i
- $\phi(o) \in \mathbb{R}^d$: Feature vector of option o ($d = 4$).
- $\theta \in \Theta$: User's preference parameter.
- θ^* : User's actual preference.
- $o_i^* = \arg\max_{o \in O_i} \ {\theta^*}^\top \phi(o)$: User's actual choice in round i

Each $\theta_j \in \{-1, -0.5, 0, 0.5, 1\}, \quad j = 1, \ldots, 4$ (departure time, duration, # stops, price).

In this case, $|\Theta| = 5^4 - 1 = 624$.

E.g., $\theta^* = (0, 0, -1, 0)$: the user only cares about fewer stops.
$o = (\text{departure time} = 0.5,\ \text{duration} = 0.3,\ \#\text{stops} = 0,\ \text{price} = 0.7) \Rightarrow$
$\phi(o) = (0.5, 0.3, 0, 0.7)^\top$.

Bayesian Assistant

User's choice model. The user selects the option maximizing reward:

$$r(o;\theta) = \theta^\top \phi(o), \qquad o^*(O,\theta) = \arg\max_{o \in O} \; r(o;\theta).$$

Bayesian update. After observing user's choice o_i^* in round i:

$$q^i(\theta) = \frac{p(o_i^* \mid \theta,\, O_i) \;\cdot\; q^{i-1}(\theta)}{\sum_{\theta' \in \Theta} p(o_i^* \mid \theta',\, O_i) \cdot q^{i-1}(\theta')},$$

where the likelihood is deterministic:

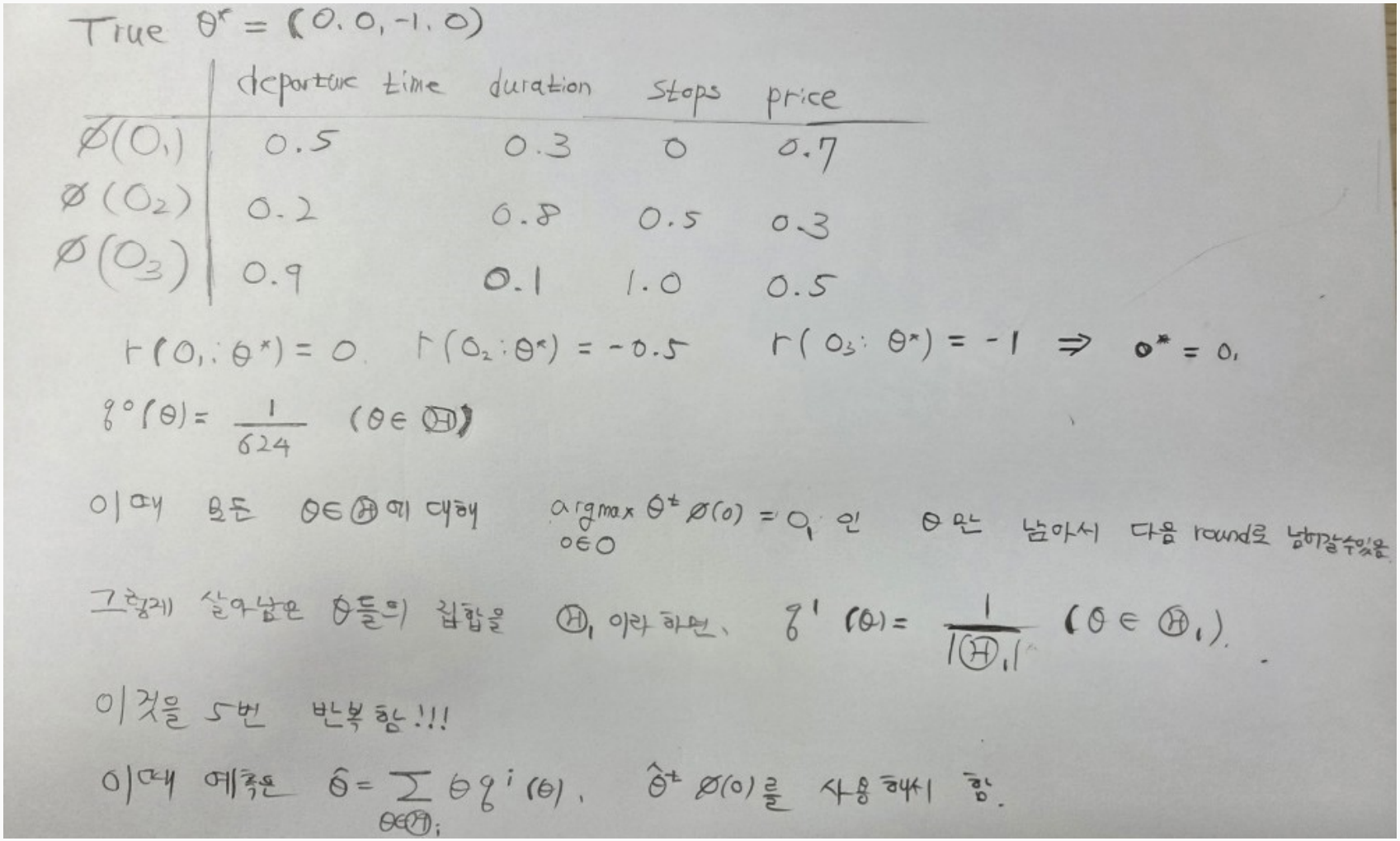
$$p(o^* \mid \theta, O) = \mathbf{1}\Big[o^*(O,\theta) = o^*\Big].$$

Inference. Use the posterior mean to inference:

$$\hat{\theta} = \mathbb{E}_{q^i}[\theta] = \sum_{\theta \in \Theta} q^i(\theta) \cdot \theta, \qquad \text{recommend } \arg\max_{o \in O_{\text{new}}} \hat{\theta}^\top \phi(o).$$

Prior: $q^0(\theta) = 1/624$ (uniform).

Example: Bayesian Assistant



Bayesian Teaching

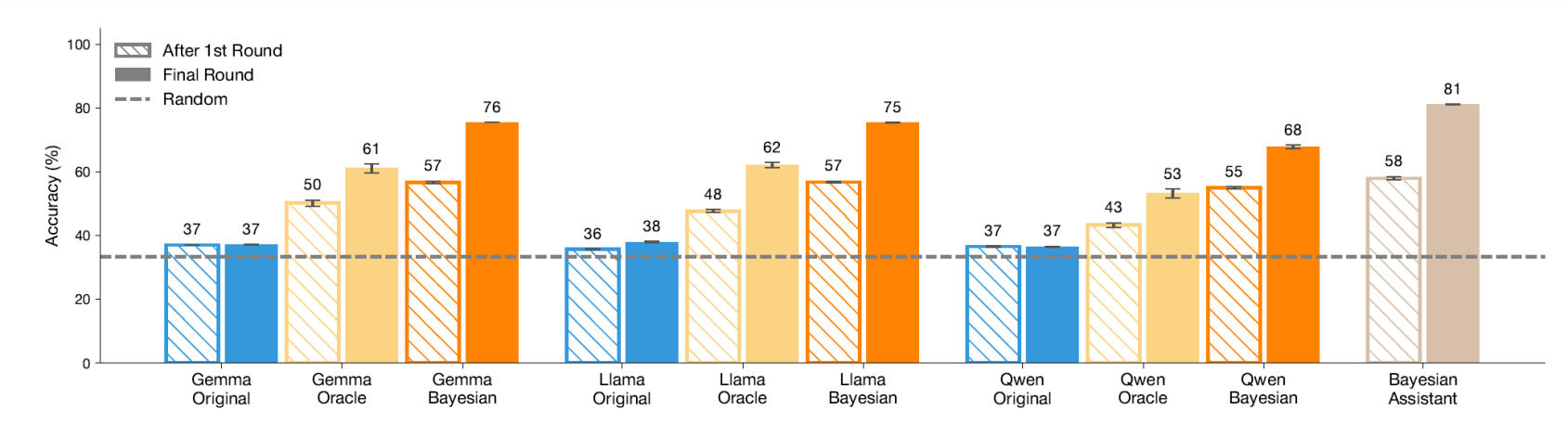
Idea: Fine-tune the LLM to mimic the Bayesian Assistant from interaction transcripts.

- Training data: for each of the 624 user preference parameters, generate 10 five-round interactions

$$\Rightarrow 624 \times 10 = 6{,}240 \text{ examples.}$$

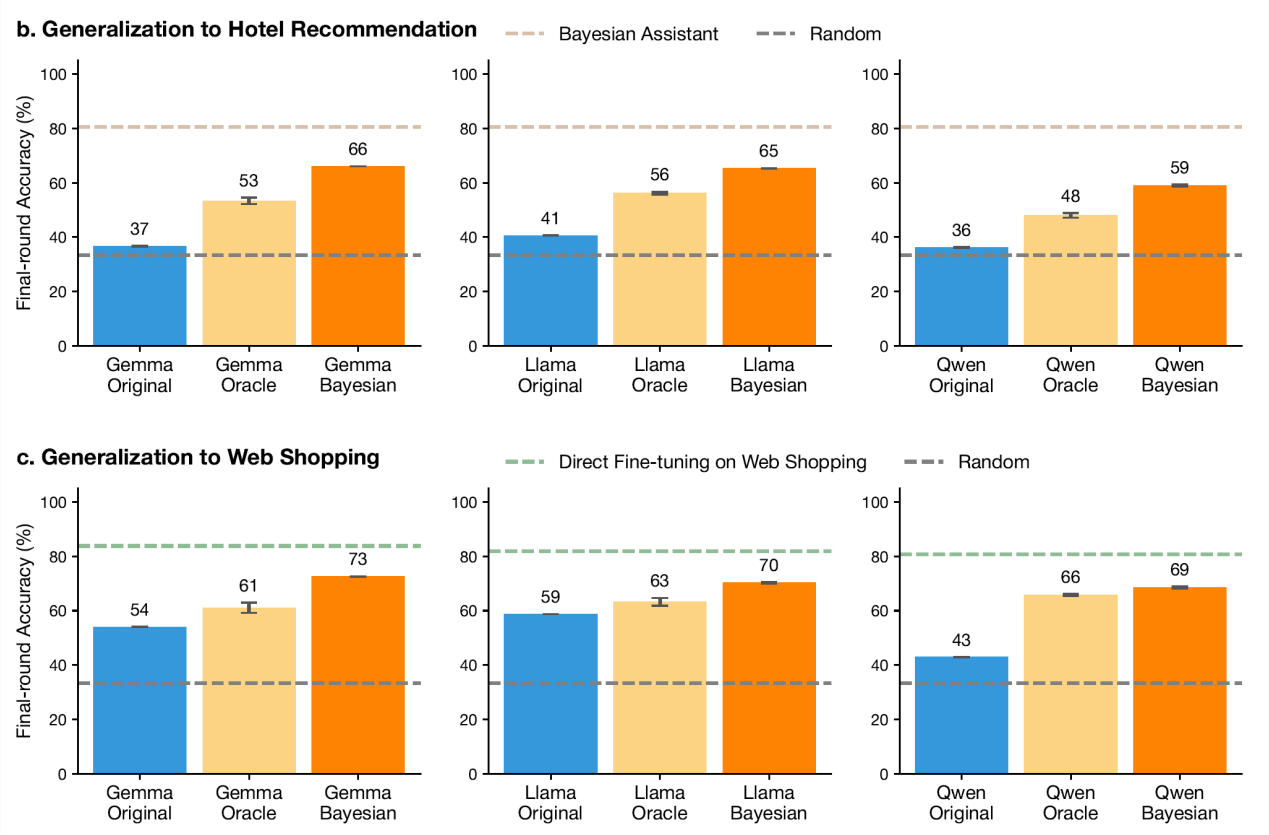
- Fine-tuned models: Gemma 2 9B, Llama 3 8B, and Qwen 2.5 7B.
- Training method: full fine-tuning with the standard language modeling objective.
- Hyperparameters: learning rate 2×10^{-6}, batch size 128, maximum sequence length 2048.

Results



- Oracle teaching: the LLM is trained on interactions in which the assistant always recommends the correct option.
- Bayesian teaching yields consistently higher accuracy than both the original LLMs and oracle-tuned models.

Results: Generalization



- The benefit of Bayesian teaching transfers beyond the flight recommendation task.
- Bayesian-tuned models achieve the best accuracy in both hotel recommendation and web shopping.