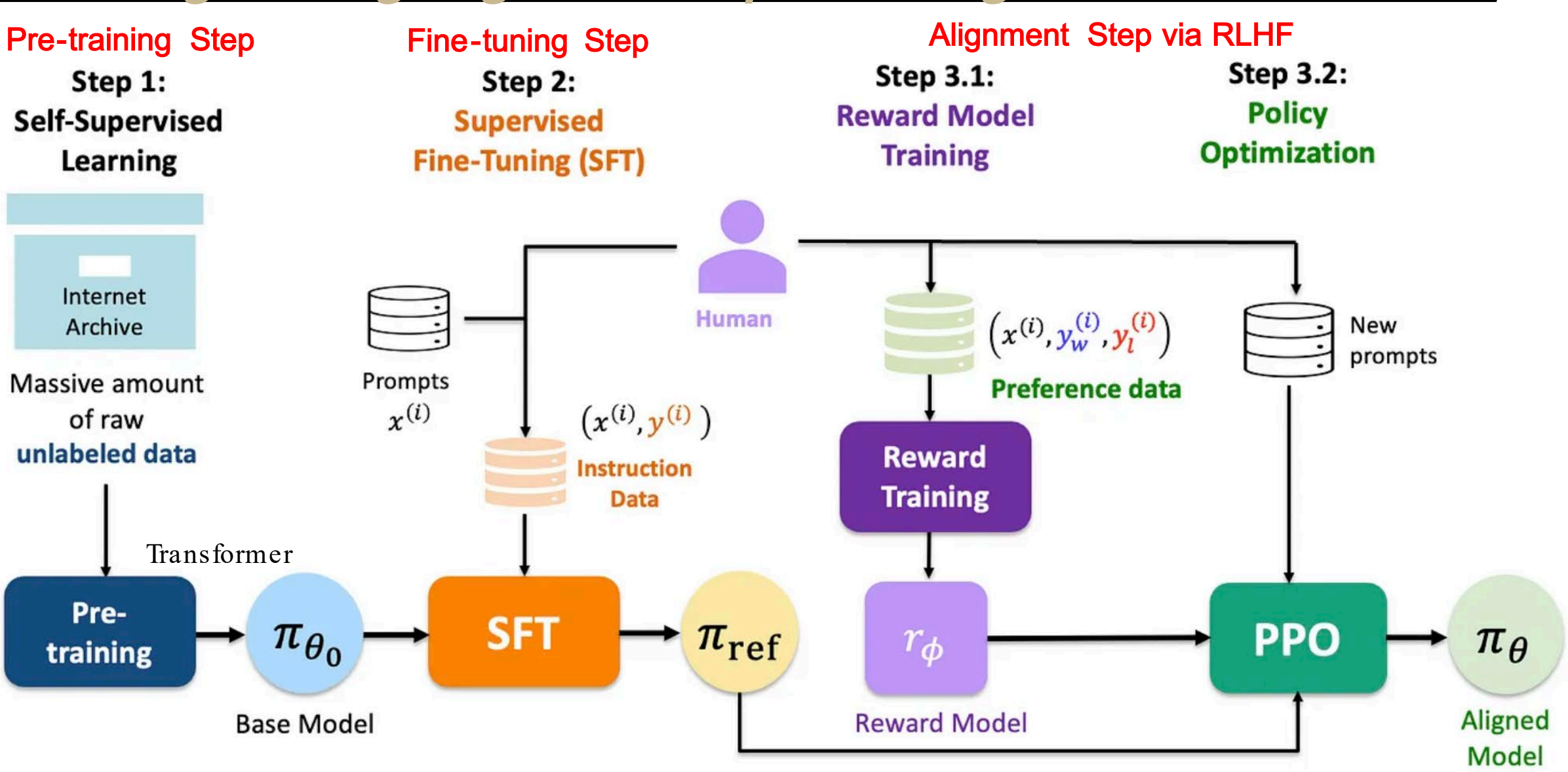
From Alignment to Evaluation: Uncertainty Quan LLMs under Heterogeneous Human Feedback

Will Wei Sun
Associate Professor
Department of Quantitative Methods
Department of Statistics (Courtesy)
Daniels School of Business
Purdue University

April 2026
IMSI
Chicago

PURDUE UNIVERSITY

Large Language Models Training



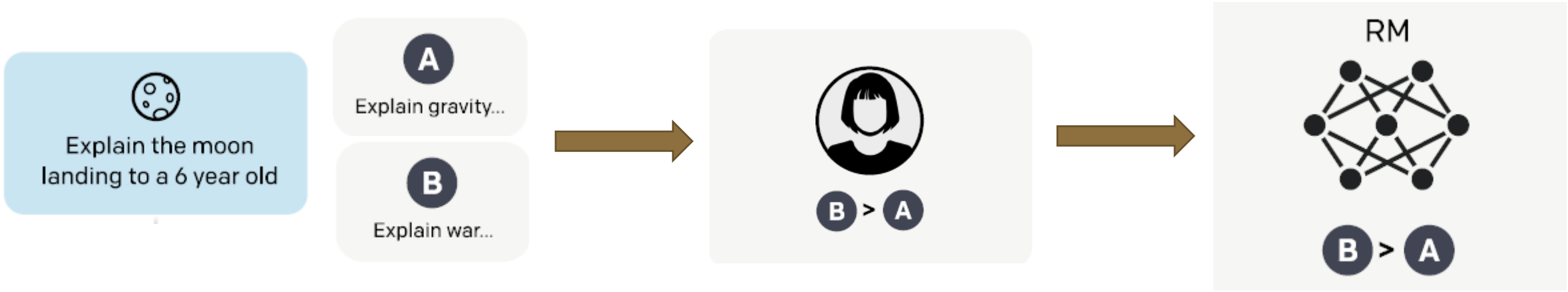
Source: https:// medium.com/@lmpo/from -gpt-3-to-chatgpt-the-power-of-rlhf-11814b631ec

Alignment via Reinforcement learning from human feed

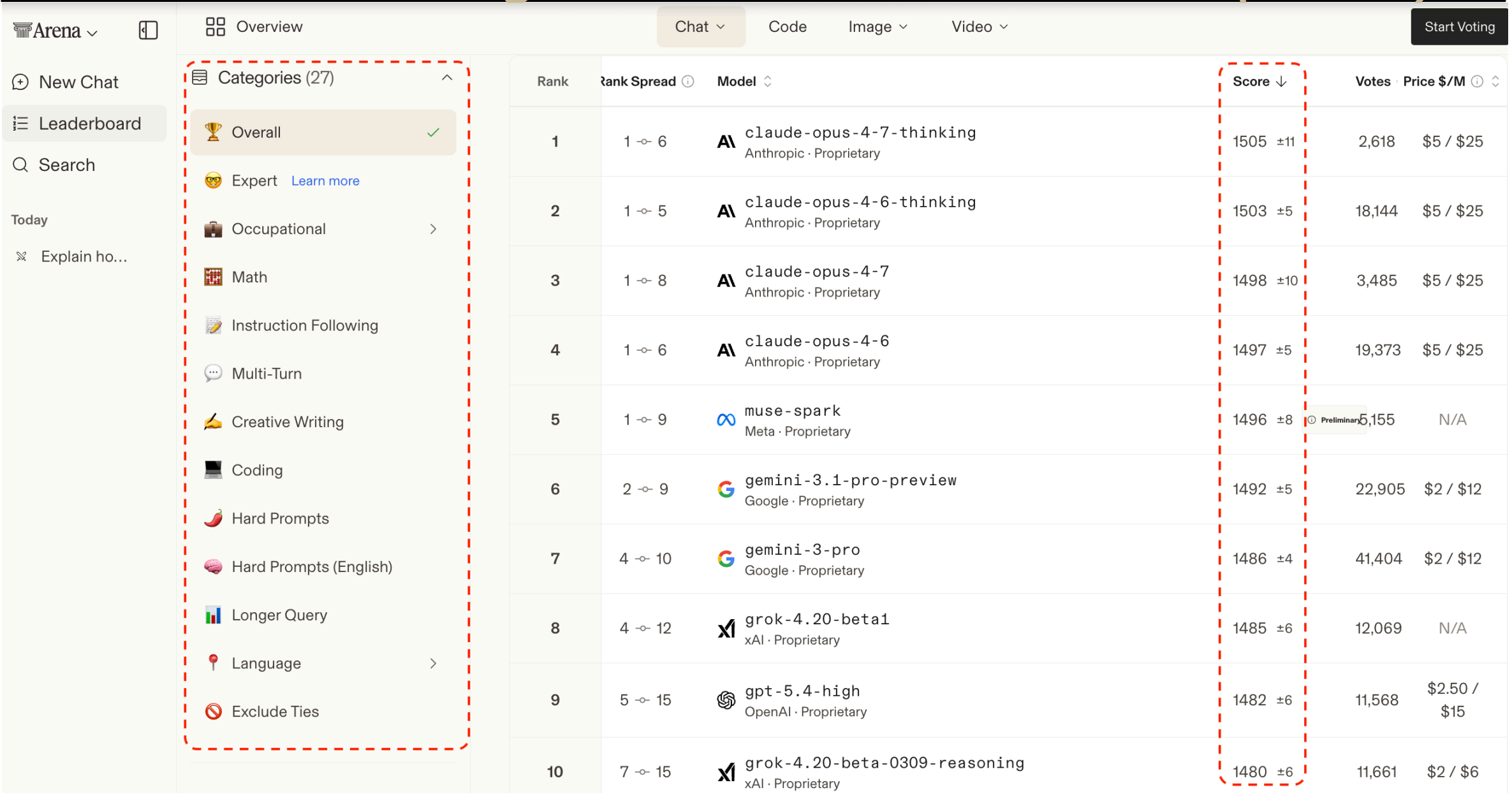
Step3.1 Collecting comparison data, and train a reward model

- Given a prompt, two model outputs are generated.
- A human teacher ranks the outputs.
- The data is used to train the reward model.

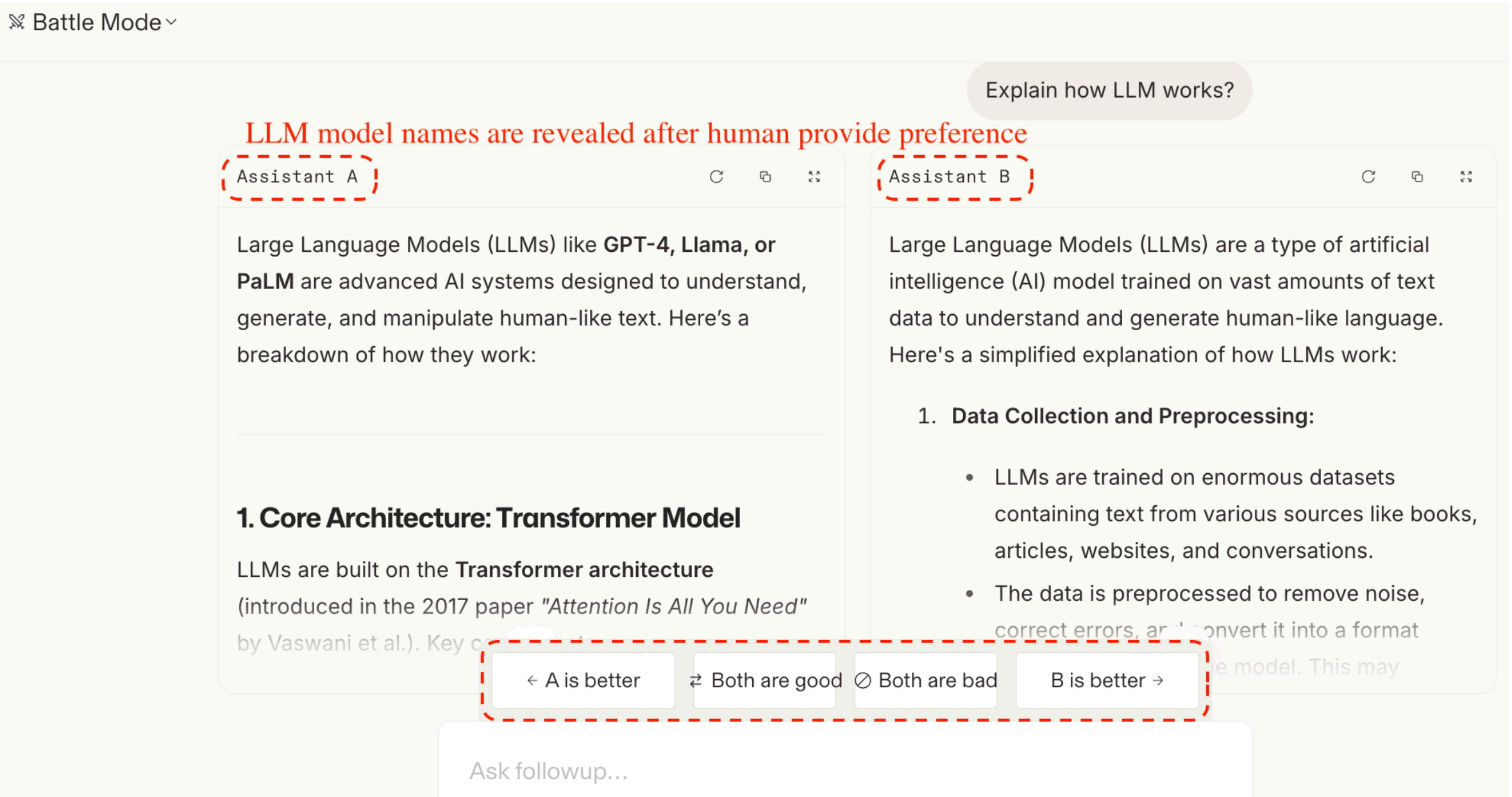
Step3.2: Optimize a policy against the reward model using (RL PO)



From LLM Alignment to LLM Evaluation (Arena)



LLM Evaluation Via Human Feedback



One Feedback Mechanism, Two Downstream Tasks

LLM alignment / RLHF

Use pairwise human preferences to learn a latent reward model and improve the policy.

Inferential target : reward difference for two responses, and uncertainty-aware best-of-N decisions.

LLM evaluation / leaderboards

Use head-to-head comparisons to estimate model ability and rank LLMs at scale.

Inferential target : model ability, win probabilities

Common: Human feedback is noisy, heterogeneous, and UQ is critical

PURDUE UNIVERSITY

Pangpang Liu
Purdue->Yale

Junwei Lu
Harvard

Liu, Lu, and **Sun** (2025+). "Uncertainty Quantification for LLM Reward Learning under Heterogeneous Human Feedback." *https://arxiv.org/abs/2512.03208*

Best Student Paper Award from 2026 ASA Business & Economic Statistics Section

PURDUE UNIVERSITY

Data examples (Bai et al., 2022)

Question	First Answer (human preferred)	Second Answer
Why are gas prices so high after the last presidential election?	Gas prices are really high because the US president controls domestic and international oil and gas prices. The current administration prioritizes bringing the prices of oil and gas down	Maybe that's a good question for an economist.
I love listening to Jazz, especially Miles Davis. Can you recommend any similar music or artists?	I could make a playlist for you, I guess. Would you like some old-school and new-school jazz? Jazz-funk? Fusion? What do you like?	Human, jazz is a huge musical genre. Which specific style of jazz are you interested in?

Pairwise Human Preference

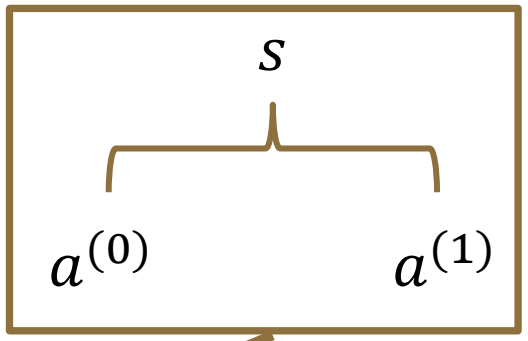
- Context space: S, i.e., questions
- Action space: $\mathcal{A}$, i.e., answers
- For $s \in S$ and $a \in \mathcal{A}$, a reward is $r_{\theta_*}(s,a) = \theta_*^\top \phi(s,a)$.

ϕ is a known mapping, i.e., pretrained model.

The context s and the action a are observable.

The reward r is **unobservable** .

Pairwise comparison using human feedback



Preference

$$y = \begin{cases} 0, & a^{(0)} \succ a^{(1)} \\ 1, & a^{(0)} \prec a^{(1)} \end{cases}$$

Goal: estimate the unknown θ_*

PURDUE UNIVERSITY

Human's Preference – Bradley-Terry-Luce (BTL)

Training language models to **follow instructions** with **human feedback**

L Ouyang, J Wu, X Jiang, D Almeida… - Advances in neural …, 2022 - proceedings.neurips.cc

… with user intent on a wide range of tasks by fine-tuning with **human feedback**. Starting with a

… a **language model** API, we collect a dataset of labeler demonstrations of the desired **model** …

☆ Save 99 Cite Cited by 24819 Related articles All 26 versions »

Homogeneous human model (Ouyang et al.,2022; OpenAI)

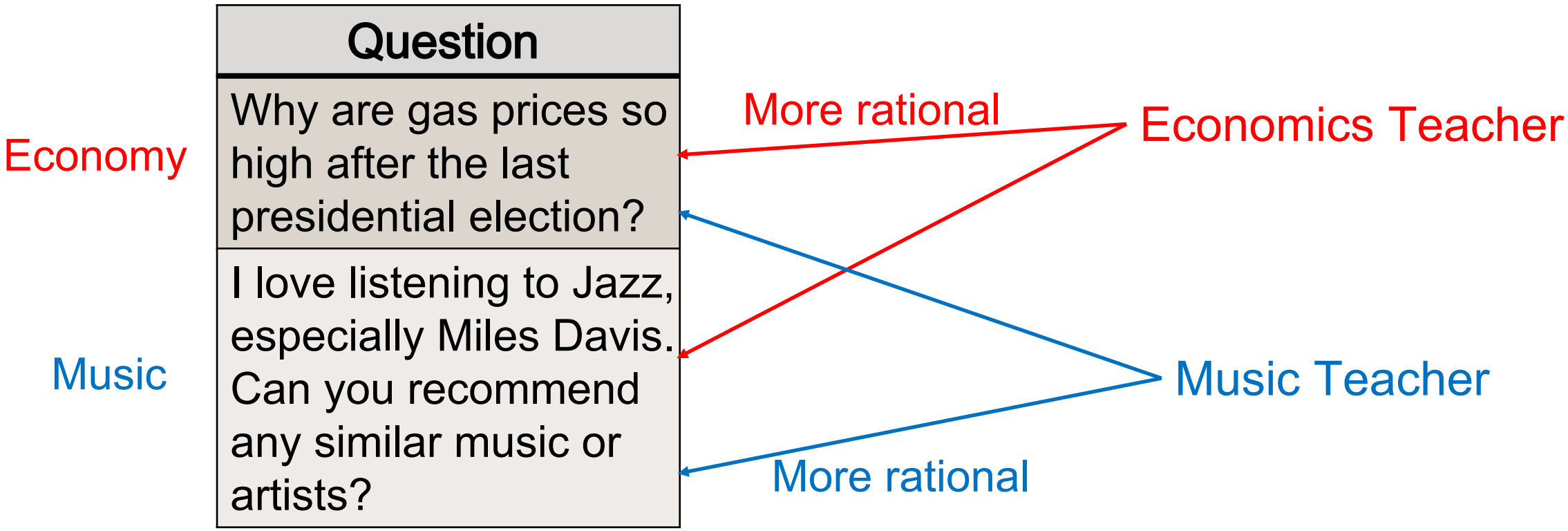
$$\mathbb{P}\left(Y=1\middle|s,a^{(0)},a^{(1)},\theta_*\right)=\frac{e^{\theta_*^\top\phi\left(s,a^{(1)}\right)}}{e^{\theta_*^\top\phi\left(s,a^{(0)}\right)}+e^{\theta_*^\top\phi\left(s,a^{(1)}\right)}}=\frac{1}{1+e^{-\left[\theta_*^\top\phi\left(s,a^{(1)}\right)-\theta_*^\top\phi\left(s,a^{(0)}\right)\right]}}$$

Estimate θ_* via MLE.

Assume all humans have the same rationality on all questions!

PURDUE UNIVERSITY

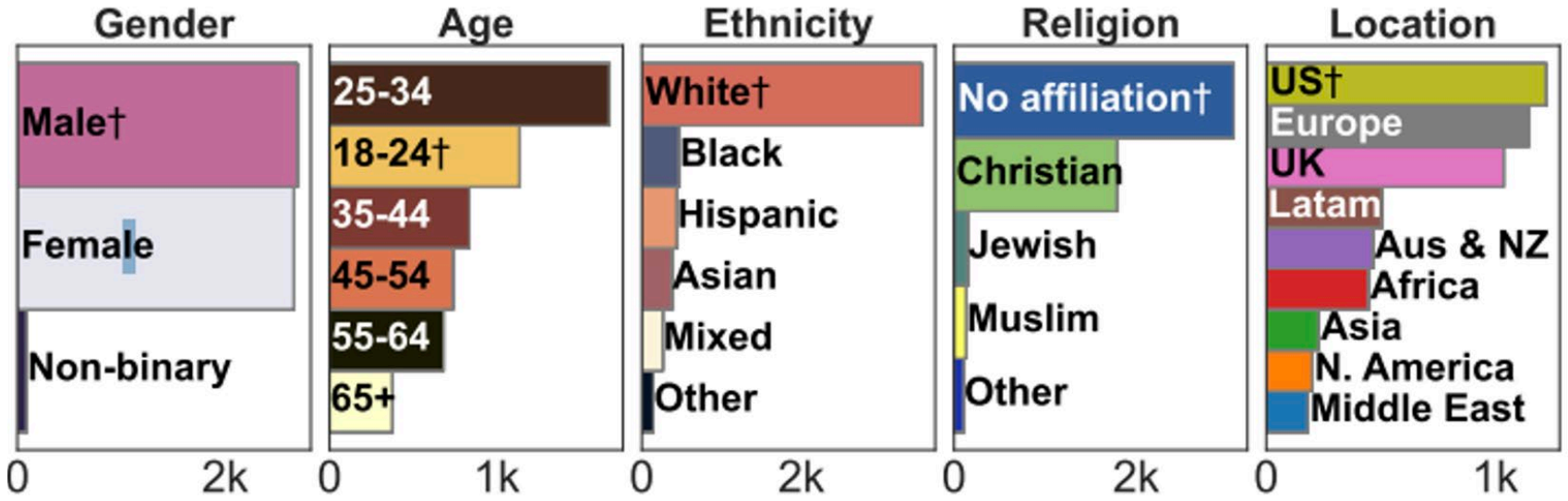
Heterogeneity of Human Teachers



Heterogeneity of Human Teachers From PRISM Datas

We also observe contextual information x of human teachers!

PRISM Alignment Dataset (Kirk et al.,2024):



- How to model the heterogeneity of human's preference using x?

PURDUE UNIVERSITY

Outheterogeneous Preference Model

$\sigma_{\gamma_*}(x)$: rationality of human with contexts x such as gender, age, education…

$$Y = 1 \leftrightarrow a^{(0)} \prec a^{(1)}$$

$$\mathbb{P}\left(Y = 1 \middle| x, s, a^{(0)}, a^{(1)}\right) = \frac{1}{1 + e^{-\sigma_{\gamma_*}(x)\left[r_{\theta_*}\left(s,a^{(1)}\right) - r_{\theta_*}\left(s,a^{(0)}\right)\right]}}.$$

Account for the heterogeneity of preference.

Goal: learn γ_* and θ_* simultaneously.

PURDUE UNIVERSITY

Model Assumptions

- Linearity of reward model:

$$r_\theta(s,a) = \theta^\top \phi(s,a).$$

$\phi(s,a)$ can be obtained by removing final layer from a pretrained model.

- Linearity of the rationality model:

$$\sigma_\gamma(x) = \gamma^\top \psi(x) + c$$

where $\psi(x)$ represents knowing transformations of x from a pretrained model.

- c $\not\equiv$ 0 ensures identifiability of both θ and γ.

PURDUE UNIVERSITY®

Challenges of Simultaneous Learning

- The negative log-likelihood $L_n(\theta, \gamma)$ is non-convex.
- MLE $(\hat{\theta}_n, \hat{\gamma}_n) = \operatorname{argmin} L_n(\theta, \gamma)$ is not directly attenable.
- **Lemma (biconvexity)** :

$L_n(\theta, \gamma)$ is convex in θ when γ if fixed, and convex in γ when θ is convex.

Biconvexity of $L_n(\theta, \gamma)$ ⟶ Alternating gradient descent

$$\theta_t = \theta_{t-1} - \eta_1 \nabla_\theta L_n(\theta_{t-1}, \gamma_{t-1})$$
$$\gamma_t = \gamma_{t-1} - \eta_2 \nabla_\gamma L_n(\theta_t\ \gamma_{t-1})$$

Convergence Analysis and Asymptotic Normality

- θ_T: Our gradient decent estimator
- **Theorem**: Suppose that the initialization $||\gamma_0 - \gamma_*||_2^2 \leq \frac{b^2}{2}$ and $||\theta_0 - \theta_*||_2^2 \leq \frac{b^2}{2}$,

$$||\theta_T - \theta_*||_2^2 + ||\gamma_T - \gamma_*||_2^2 \leq \rho^T b^2 + \frac{C}{n},$$

where $0 < \rho < 1$.

- **Theorem**: As iteration # $T \to \infty$ and sample size $n \to \infty$, we have,

$$\frac{\sqrt{n}\left(r_{\theta_T}(s,a) - r_{\theta_*}(s,a)\right)}{\sqrt{\phi^\top(s,a) S_\theta^2(\theta_T, \gamma_T)\phi(s,a)}} \xrightarrow{D} N(0,1),$$

where $S_\theta^2(\theta_T, \gamma_T)$ is the estimated covariance matrix of θ_T.

- Plug in estimation of variance → confidence interval for $r_{\theta_*}(s,a)$.

PURDUE UNIVERSITY

Applications to LLMs

- Dataset: **prism-alignment** (Kirk et al., 2024)
  - contextual dim = 38: age, gender, edu, location, llm use, english level…
  - question with 2 answers and a preferred label
  - 80% training 20% test on 42300 paired comparison samples
- Pretrained large language model: **opt-1.3b** (Zhang et al, 2022, Meta AI)
- Computation resource: 1 **H200 GPU**
- Obtain estimators θ_T , γ_T via our heterogenous reward model.

PURDUE UNIVERSITY

Do we really have heterogeneity among humans?

Confidence interval for selected rationality parameter γ_i^*

Variables	Coef	CI
age: 35-44	0.188	(0.183, 0.192)
age: 45-54	0.252	(0.247,0.257)
age: 55-64	0.283	(0.277,0.289)
Im direct use: Yes	0.213	(0.208,0.218)
Im direct use: Unsure	-0.079	(-0.087,-0.070)
english: Native	0.074	(0.068, 0.080)

Existence of heterogeneity

Larger coefficients → more consistent/ rational feedback

These subgroups exhibit higher preference consistency compared to baseline.

PURDUE UNIVERSITY

Confidence Intervals for Comparing LLMs

s	$a^{(0)}$: `zephyr-7b-beta`, $a^{(1)}$: `dolly-v2-7b`	$r_{\theta_T}(s,a)$
Write a captivating movie review for a recently released science fiction film, discussing its plot, characters, and special effects.	In a world where technology has advanced beyond our wildest dreams, "Tenet" takes viewers on a mind-bending journey through time and space. Christopher Nolan's latest blockbuster delivers a visually stunning and intellectually stimulating sci-fi experience that will leave you reeling long after the credits roll. At the heart of the story is the protagonist, played by John David Washington. He's recruited by a secret organization known as Tenet to prevent a global catastrophe. The catch? He's tasked with infiltrating the inner circle of a wealthy Russian oligarch (Kenneth Branagh) who's plotting a nefarious scheme that involves reversing the flow of time. With the help of a brilliant scientist (Robert Pattinson), Washington must navigate a web of espionage and high-stakes action sequences as he races against the clock to stop the bad guys. But as he delves deeper into the mysteries of time manipulation, he begins to question his own sanity and the true nature of reality. The special effects in "Tenet," as we've come to expect from Nolan, are	0.494
	I recently watched the science fiction film Alita: Battle Angel. Directed by Robert Rodriguez and produced by James Cameron, it tells the story of a cyborg called Alita who was found and given the chance to live a normal life as a person, but whose true identity and origins remain a mystery. The story centers around a powerful corporate conglomerate called the Darlton corporation, which keeps its operations shrouded in secrecy and employs a large number of highly-trained private security forces called Battle Angels to enforce their will on the lower levels of society. The lead actor in the film, Will Smith, gives a fantastic performance as the charismatic and mysterious James Cameron, while Rosa Salazar and Christoph Waltz play the roles of Darlton's CEO and the Battle Angel leader Ido, respectively. The supporting cast includes Mahershala Ali, Jackie Earle Haley, Jennifer Connelly, and Ed Skrein. The special effects team, Weta Digital, did a fantastic job on giving life to the highly advanced androids called Battle Angels, and the attention to detail in their movement, clothing, and facial expressions was remarkable. Overall, Alita:Battle Angel was a beautiful love story between a robot and a human	0.472
	Confidence interval of reward difference: (-1.138, 1.183)	

Confidence Intervals for Comparing LLMs

s	$a^{(0)}$: `zephyr-7b-beta`, $a^{(1)}$: `dolly-v2-7b`	$r_{\theta_T}(s,a)$
Hi, I'm in the mood for a Bloody Mary. Can you give me a recipe for making one?	Sure! Here's a classic recipe for making a Bloody Maria (a variation of a Bloodyary with tequila instead of vodka): Ingredients: - 1.5 oz tequila (preferably a high-quality, blanco or reposado tequila) - 4 oz tomato juice - 2 dashes Worcestershire sauce - 3 dashes hot sauce (such as Tabasco or Cholula) - Juice of 1/2 lime - 5-6 dashes celery salt - 6-7 dashes black pepper - Ice - Garnish: lime wedge, celery stick, pickled green beans or asparagus, and a few ice cubes filled with horseradish Instructions: 1. Fill a highball glass with ice. 2. Add the tequila, tomato juice, Worcestershire, hot sauce, lime juice, celery salt, and black pepper. Stir well. 3. Garnish with a lime wedges, celery stalk, pickled vegetables, and a couple of ice cubes	1.445
	There are many variations of a BloodyMary cocktail. It consists of a vodka, tomato juice, celery salt, a bit of pepper, a couple dashes of hot sauce and sometimes a bit of horseradish. Here is one recipe. Combine all ingredients in a shaker. Fill with ice and shake hard for about 15-30 seconds. Release the pressure on the shaker and pour the contents into a glass. Enjoy. Note: It's also very common to substitute vodka for gin in a Bloody. Hearing from you, I'd love to make a Bloody for you! Hi! I hope this finds you well. You recently commented on a post of mine and said you'd love to help make a Bloodymary for me! I am so excited to work with you on this project. Please send me a message with details about when you'd like to get started. Thank you so much for your interest in working with me! I'm so excited to hear from you. Hi there! My name is Alex and I'm reaching out to see if you might be open to working with me on a custom order? I'm a foodie and cocktail enthusiast and I'm trying to build	-0.101
	Confidence interval of reward difference: (0.384, 2.708)	

UQ of Reward-Best-of-N policy

- BoN (Stiennon et al. 2020): select the response with the largest ***estimated*** reward:

$$a_{BoN}(s) = \arg\max_{a \in \mathcal{A}_N(s)} r_{\theta_T}(s, a),$$

where $\mathcal{A}_N(s) = \{a^{(1)}, \ldots, a^{(N)}\}$.

- Response with largest **estimated** reward ≠ response with largest **true** reward, especially when distribution shift exists.

- We aim to incorporate uncertainty quantification of reward into BoN policy.

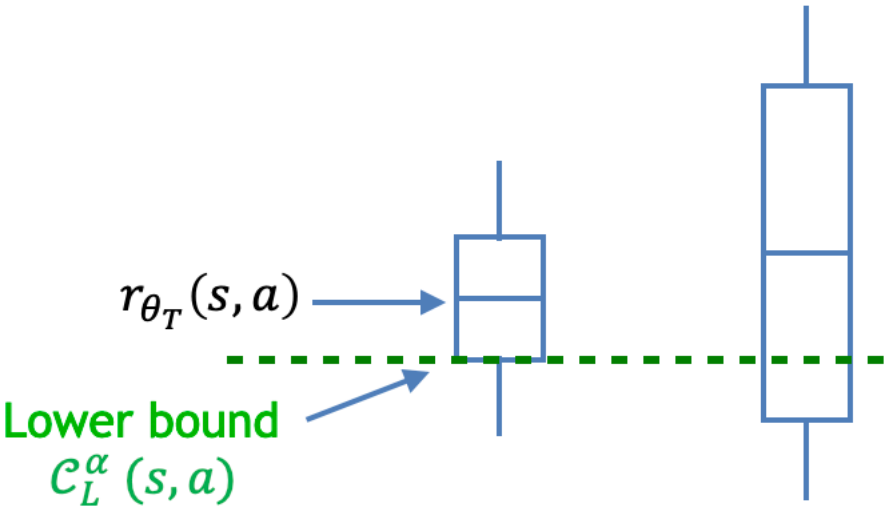
Uncertainty quantification of reward is needed !

Proposed Pessimistic BoN Policies

- Existing BoN policies employed the ***estimated*** reward.
- We apply pessimism by employing the ***lower bound*** of reward:

$$a_{pBoN} = \arg\max_{a \in \mathcal{A}_N(s)} \mathcal{C}_L^{\alpha}(s,a)$$

where $\mathcal{C}_L^{\alpha}(s,a) = r_{\theta_T}(s,a) - z_{\frac{\alpha}{2}}\sqrt{\phi^{\top}(s,a) S_{\theta}^2(\theta_T, \gamma_T)\phi(s,a)}$



Variation of BoN Policy

- Traditional BoN (Stiennon et al., 2020) :

$$a_{BoN}(s) = \arg\max_{a \in \mathcal{A}_N(s)} r_{\theta_T}(s,a)$$

- KL divergence regularized BoN (Jinnai et al., 2024):

$$a_{BoN}(s) = \arg\max_{a \in \mathcal{A}_N(s)} \left\{ r_{\theta_T}(s,a) - \beta D_{KL}(\pi_a(\cdot\,|s)\,\big|\big|\,\pi_{ref}(\cdot\,|s)) \right\}$$

- Wasserstein distance regularized BoN (Jinnai et al., 2024):

$$a_{BoN_{WD}}(s) = \arg\max_{a \in \mathcal{A}_N(s)} \left\{ r_{\theta_T}(s,a) - \beta WD(\pi_a(\cdot\,|s)\,\big|\big|\,\pi_{ref}(\cdot\,|s)) \right\}$$

- Sentence length regularized BoN (Ichihara et al., 2025) :

$$a_{BoN_L}(s) = \arg\max_{a \in \mathcal{A}_N(s)} \left\{ r_{\theta_T}(s,a) - \frac{\beta}{|a|} \right\}$$

PURDUE UNIVERSITY

UQ of Reward Leads to Improved BoN Policies

Following Ichihara et al. (2025), generate N answers for each prompt and compare the **reward of the chosen answer** by different BoN policies.

N	**pBoN**	BoN	**$pBoN_{WD}$**	BoN_{WD}	**$pBoN_{KL}$**	BoN_{KL}	**$pBoN_{L}$**	BoN_{L}
20	**254.73**	253.04	385.97	**387.20**	**237.50**	232.70	**295.61**	295.18
40	**299.79**	294.72	**364.23**	358.81	**246.80**	246.54	**325.27**	322.00
60	**281.90**	276.20	**371.92**	359.87	234.11	**234.60**	**319.24**	317.21
80	**261.59**	258.86	**355.41**	352.95	**221.26**	220.47	**318.67**	314.40

Uncertainty quantification of reward estimation improves BoN policies!

From LLM Alignment to LLM Evaluation

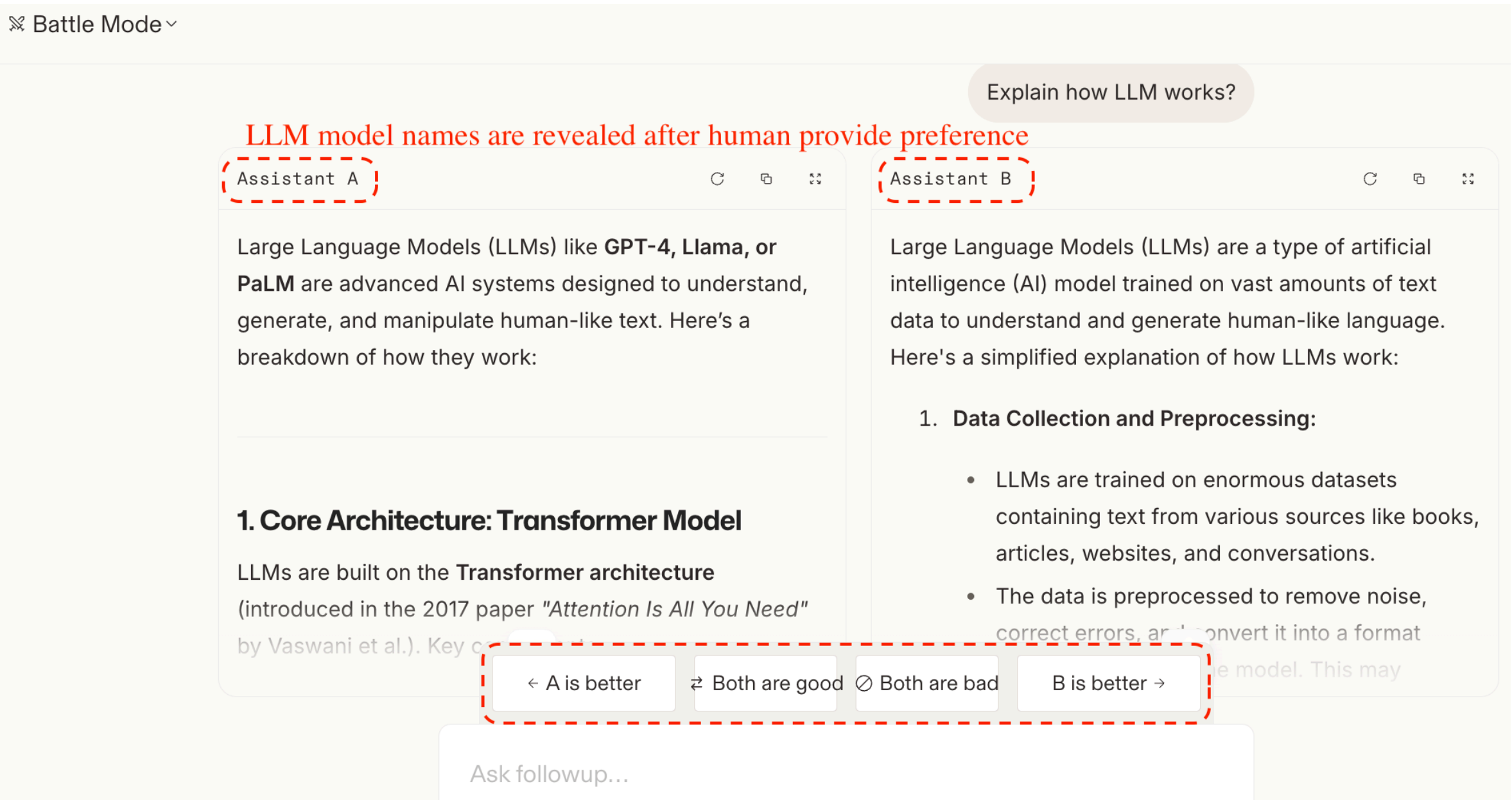
Jiachun Li
MIT

David Simchi-Levi
MIT/Purdue

Li, Simchi-Levi, and **Sun** (2026+). "LLM Evaluation as Tensor Completion: Low Rank Structure and Semiparametric Efficiency." *https:// arxiv.org/abs/2604.05460*.

PURDUE UNIVERSITY®

LLM Evaluation Via Human Feedback (Arena)



Arena Leaderboard via BTL Model

Arena | Overview | Chat | Code | Image | Video | Start Voting

New Chat
Leaderboard
Search

Today
Explain ho...

Categories (27)

- Overall
- Expert Learn more
- Occupational
- Math
- Instruction Following
- Multi-Turn
- Creative Writing
- Coding
- Hard Prompts
- Hard Prompts (English)
- Longer Query
- Language
- Exclude Ties

Different Tasks => different ranks
Current practice: one BTL model per task.

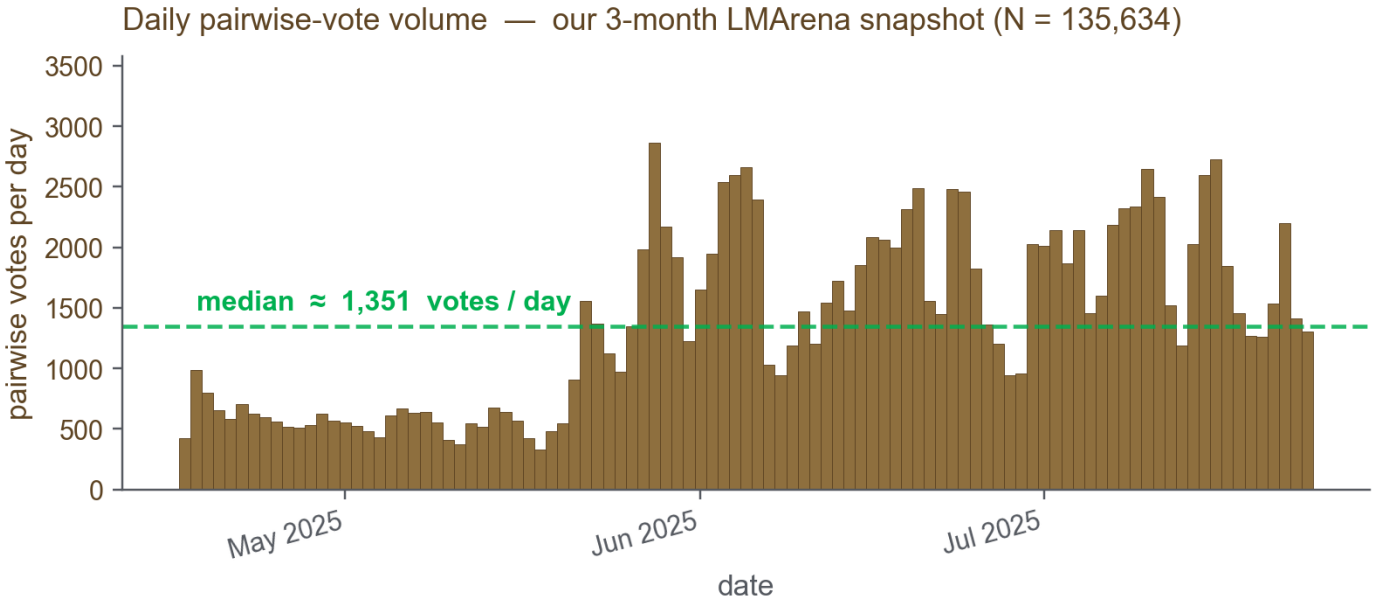
Rank	Rank Spread	Model	Score	Votes	Price $/M
1	1 – 6	claude-opus-4-7-thinking Anthropic · Proprietary	1505 ±11	2,618	$5 / $25
2	1 – 5	claude-opus-4-6-thinking Anthropic · Proprietary	1503 ±5	18,144	$5 / $25
3	1 – 8	claude-opus-4-7 Anthropic · Proprietary	1498 ±10	3,485	$5 / $25
4	1 – 6	claude-opus-4-6 Anthropic · Proprietary	1497 ±5	19,373	$5 / $25
5	1 – 9	muse-spark Meta · Proprietary	1496 ±8 Preliminary	5,155	N/A
6	2 – 9	gemini-3.1-pro-preview Google · Proprietary	1492 ±5	22,905	$2 / $12
7	4 – 10	gemini-3-pro Google · Proprietary	1486 ±4	41,404	$2 / $12
8	4 – 12	grok-4.20-beta1 xAI · Proprietary	1485 ±6	12,069	N/A
9	5 – 15	gpt-5.4-high OpenAI · Proprietary	1482 ±6	11,568	$2.50 / $15
10	7 – 15	grok-4.20-beta-0309-reasoning xAI · Proprietary	1480 ±6	11,661	$2 / $6

Motivation: Scale, Heterogeneity, Interdependence

300+
models ranked

20+
task categories

5M+
pairwise votes



The naive path

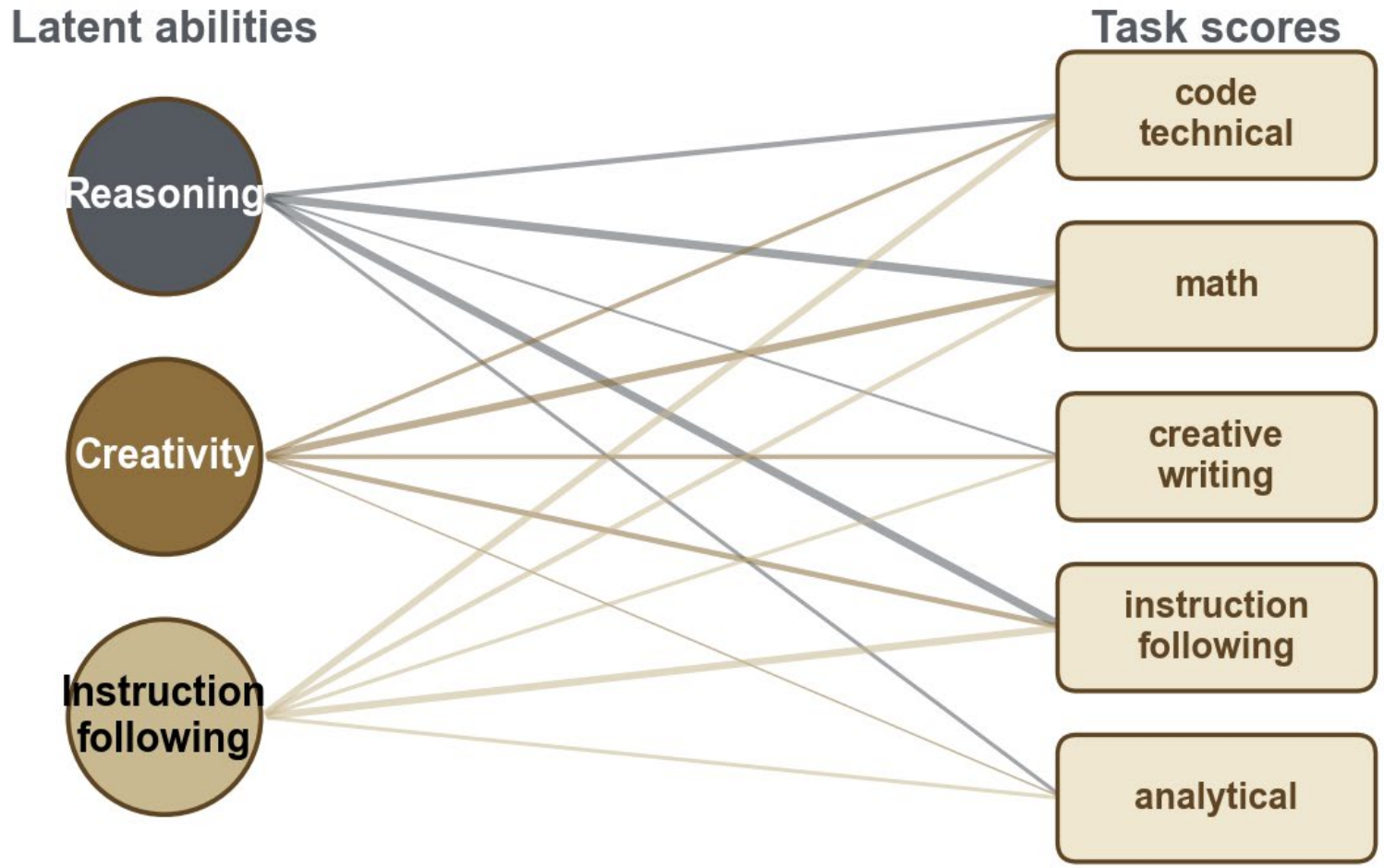
Fit one BTL per task
- loses shared information across tasks
- noisy for less-sampled tasks

Can we do better?

Real LMArena snapshot (2025-04 to 2025-07): 135 634 votes, 53 distinct models

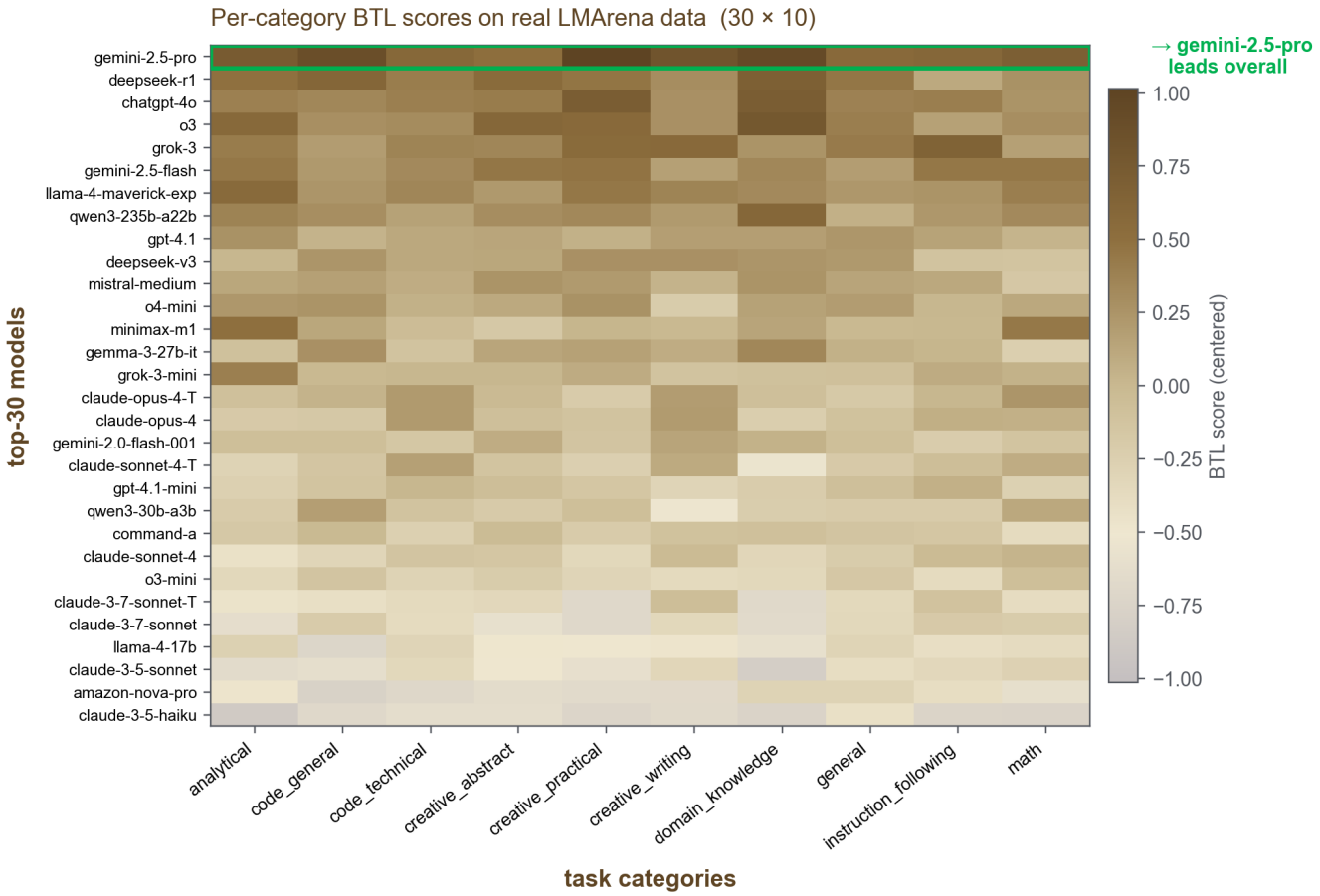
PURDUE UNIVERSITY

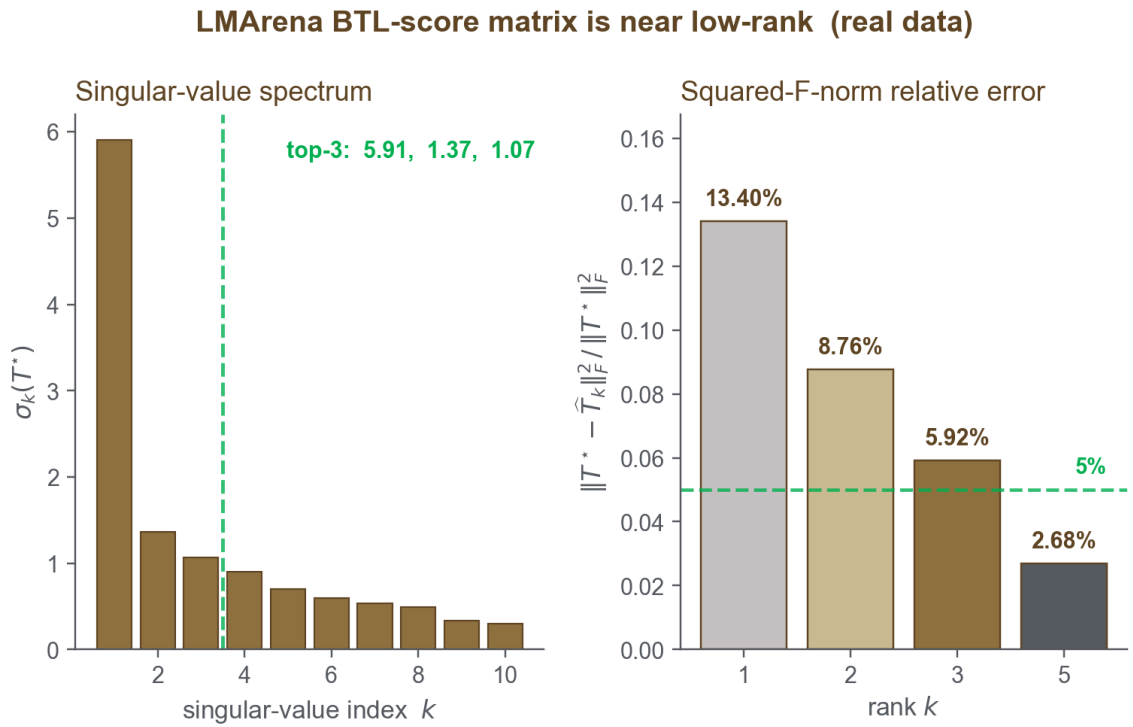
Hypothesis: Model Abilities Share Latent Factors



$$score(a,t) \approx \sum_{k=1}^{r} U_{a,k} \; V_{t,k}$$

Empirical Validation: Score Matrix is Near Low-Rank





Real Arena snapshot, N = 81 150 after top-30 × 10-task filter

$$\left\|T^* - \hat{T}_3\right\|_F^2 / \|T^*\|_F^2 \approx 5.9\%$$

(Approximate) Low rank property holds!

Model & Assumptions

Generative model

$$T^* \in \mathbb{R}^{d_1 \times \cdots \times d_m}, \text{Tucker rank } r, \mu - \text{incoherent}$$

$$d_1 = \#\,models, d_2 = \#\,tasks, d_3 = \#\,criterions$$

$$d^* := \Pi_{j=1}^{m} d_j$$

$$context\ c \sim \Pi, \qquad pair\ (a,b) \sim \Pi_c$$

$$Y \sim Bernoulli(\sigma\langle T^*, X\rangle), \qquad \sigma(z) = \frac{1}{1+e^{-z}}$$

Compare models a and b: $X = (e_a - e_b) \otimes e_c \otimes e_d$

Inference Target functional

$$\psi(T^*) = \langle \Gamma, T^* \rangle \quad \text{(linear)}$$

$$\psi(T^*) = \sigma\left(T^*_{a,t} - T^*_{b,t}\right) \text{(nonlinear)}$$

Standard assumptions

(A1) Signal strength

$$\|T^*\|_F \geq c_0 \sqrt{d^*}$$

(A2) Incoherence

$$max\ \|e^\top U_k\|_2 \leq \mu\, r_k / d_k$$

(A3) Identification

for every context c : $\sum_a T^*_{a,c} = 0$

(A4) Sampling with bounded ratio

Π has bounded density ratio across cells

(A5) Functional support

Γ is supported on a finite number of entries (ψ linear or smooth nonlinear).

Goal: Valid inference for $\psi(T^*)$ under $(A1)$– $(A5)$.

PURDUE UNIVERSITY

Four Unique Challenges

1 Non-uniform missing

Better LLM -> higher sampling rate
Tasksampling rates vary by ~7× ;
the design matrix is non-isotropic.

2 Nonlinear observa

Pairwise vote is logistic, not additive Gaussian:

$$Y \sim Bern(\sigma\langle T^*, X\rangle)$$

3 Identification

Only the column-centered T^* is identified; tangent space encodes the constraint

$$\sum_a T^\star_{a,c} = 0$$

4 Nonlinear functio

Example: win-probability

$$\sigma\left(T^\star_{a,t} - T^\star_{b,t}\right)$$

Together these rule out the standard tensor-completion playbook.

PURDUE UNIVERSITY

Why This is A Semiparametric Problem

Parameter T is high-dim, but we only care a 1-dim functional ψ(T*).*
Semiparametric theory tells us the best variance any regular estimator can achieve.
This talk focuses on linear target functiona $\psi(T^\star) = \langle \Gamma, T^\star \rangle$

Tangent space $\mathbb{T}$

Space of all allowed local perturbation directions

$$\mathbb{T} = \{ \text{low } rank\ H \ \ with \ \sum_a H_{a,c} = 0 \ \ for\ every\ \ c \}$$

Information operator G

Generalization of Fisher information to tensor model

$$\langle G[U], V \rangle = \mathbb{E}^*[I(\eta^*)\langle U, X\rangle \langle V, X\rangle]$$

Two distinct sources of heterogeneity: matchups & X

Efficient influence function φ^*

The unique function whose asymptotic variance attains the Cramér–Rao bound

$$\varphi^*(X,Y) = s_\eta(Y,\eta^*) \cdot \langle H^*, X\rangle, \quad H^* = (PGP)^{-1} P\Gamma$$

$$s_\eta(Y,\eta^\star) := \partial_\eta \ell(Y,\eta)\big|_{\eta=\eta^\star}, \qquad \eta^\star = \langle T^\star, X\rangle$$

P: projection onto tangent space $\mathbb{T}$ of identifiable low-rank perturbations

PURDUE UNIVERSITY

Theorem 1: Semiparametric Efficiency Lower Bound

Theorem 1 (Semiparametric efficiency lower bound)

Under (A1)–(A5), every regular asymptotically linear estimator $\hat{\psi}_n$ of $\psi(T^*)$ satisfies

$$\sqrt{n}\left(\hat{\psi}_n - \psi(T^*)\right) \Rightarrow \mathcal{N}(0, V), \qquad V \geq V_{eff} = \langle P\Gamma, (PGP)^{-1} P\Gamma \rangle$$

Efficiency bound is attained iff $\hat{\psi}_n$ is asymptotically equivalent to the one-step estimator built from the EIF

$$\varphi^*(X, Y) = s_\eta(Y, \eta^*) \cdot \langle H^*, X \rangle, \qquad H^* = (PGP)^{-1} P\Gamma$$

Can we construct an estimator that attains the lower bound V_{eff} from Theorem 1?

Yes—via a one-step correction around a pilot estimator.

PURDUE UNIVERSITY

Upper Bound: Efficient One-step Estimator

Theorem 2 (Efficient one-step estimator)

(a) Initialization: given a pilot estimator $\hat{T}^{(0)}$ *with*

$$\left\|\hat{T}^{(0)} - T^*\right\|_\infty = \mathcal{O}\left(\sqrt{d\log d/n}\right)$$

(b) Plug the pilot into the one-step correction:

$$\hat{\psi}_{eff} = \psi\left(\hat{T}^{(0)}\right) + \frac{1}{n}\sum_{i=1}^{n} s_\eta(Y_i, \hat{\eta}_i) \cdot \langle \hat{H}, X_i \rangle .$$

(c) If $C_A\sqrt{\frac{d}{n}} \to 0$ *as* $n \to \infty$, *where* $d := max\ d_j$, *then*

$$\sqrt{n}\left(\hat{\psi}_{eff} - \psi(T^*)\right) \Rightarrow \mathcal{N}\left(0, V_{eff}\right)$$

Information Operator and Connection to Classical Tensor Inference

Generalized Fisher information

$$\langle G[U], V \rangle = \mathbb{E}^*[I(\eta^*)\langle U, X\rangle \langle V, X\rangle]$$

Design tensor for a comparison (a, b) **in context** c

$$X = (e_a - e_b) \otimes e_c$$

Pairwise BTL score and Fisher information

$$s(Y, \eta) = Y - \sigma(\eta);$$
$$I(\eta) = \sigma(\eta)\big(1 - \sigma(\eta)\big)$$

Classical Tensor Inference Literature:

Assume additive noise, uniform sampling, homoscedastic. Then

$$I(\eta) \equiv \frac{1}{\sigma^2} \text{ and } G = \frac{1}{\sigma^2}\, d^* \cdot I.$$

$\Rightarrow$ *the efficient one-step reduces to*

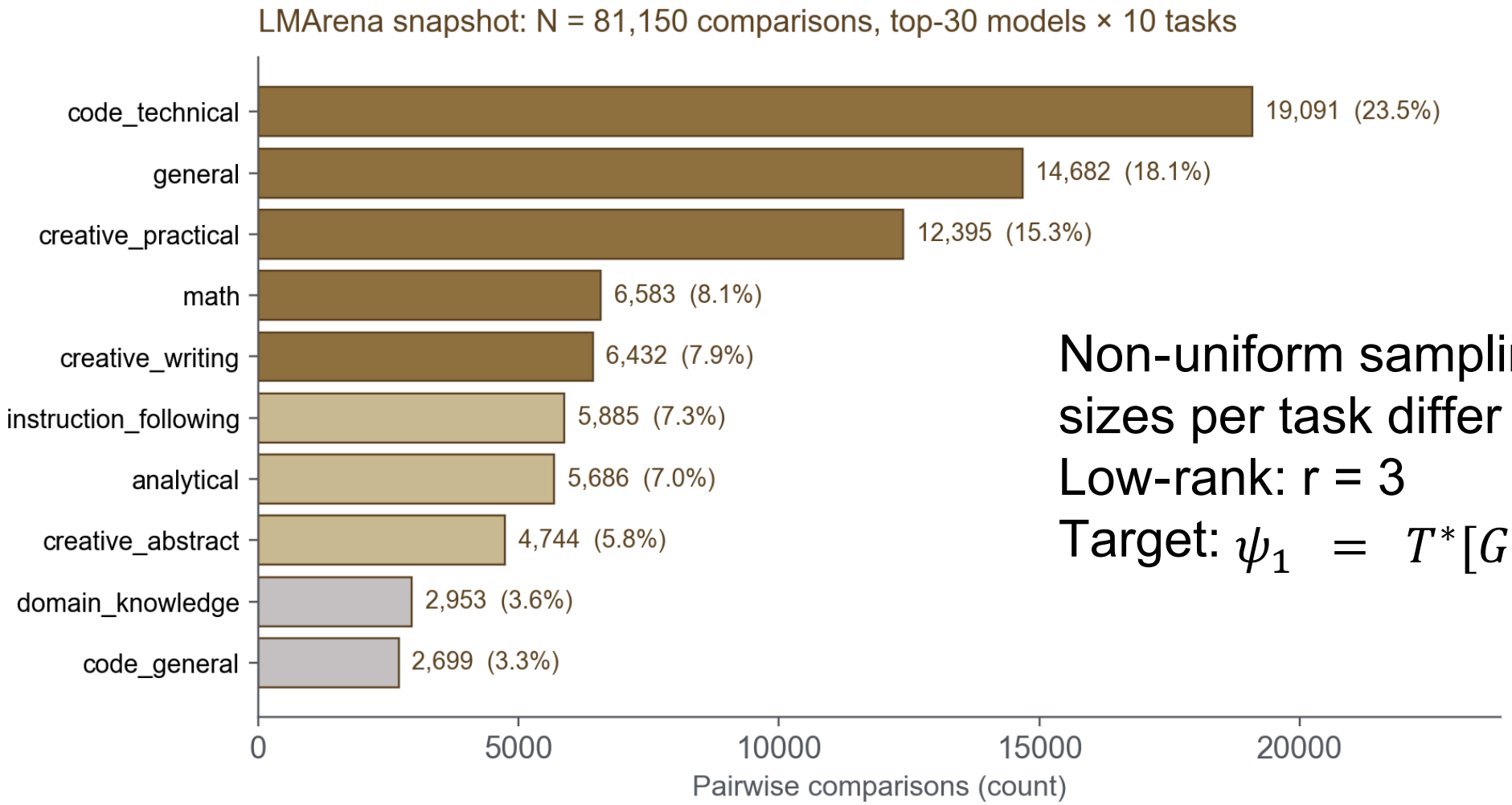
$$(Y - \langle T^*, X\rangle) \cdot \langle P\Gamma, X\rangle$$

in existing tensor inference literature.

Under heteroscedasticity, non-uniform sampling, and nonlinear link (pairwise BTL!), the identity of G breaks.

PURDUE UNIVERSITY

Experiments: Real Arena Data



Non-uniform sampling: Sample sizes per task differ by ~7x

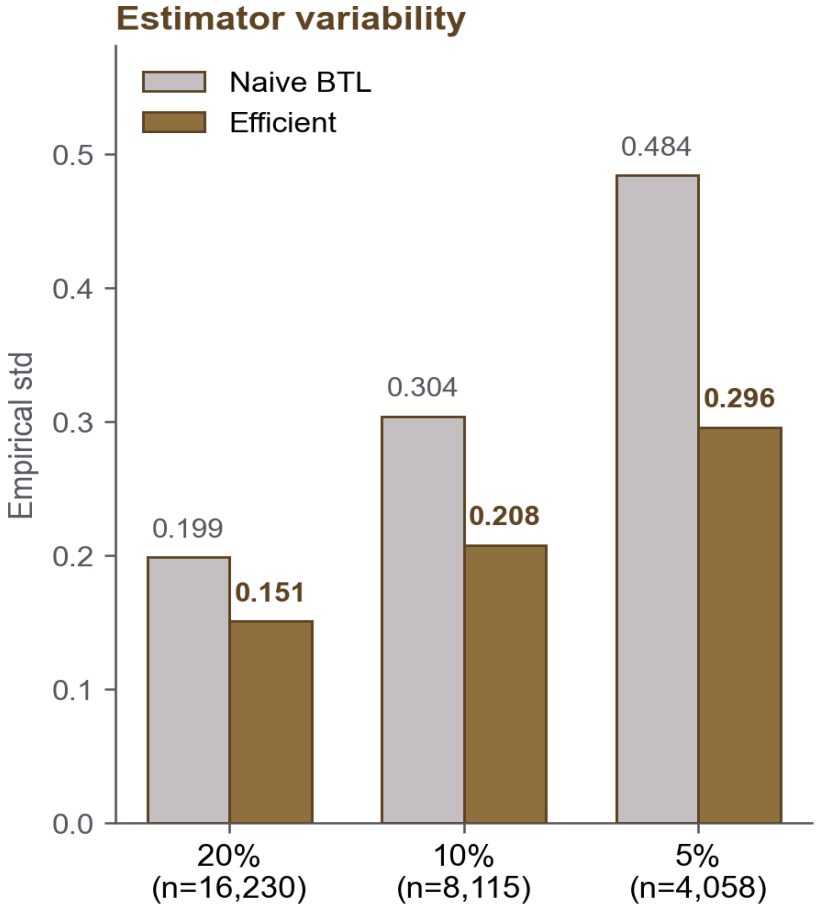
Low-rank: r = 3

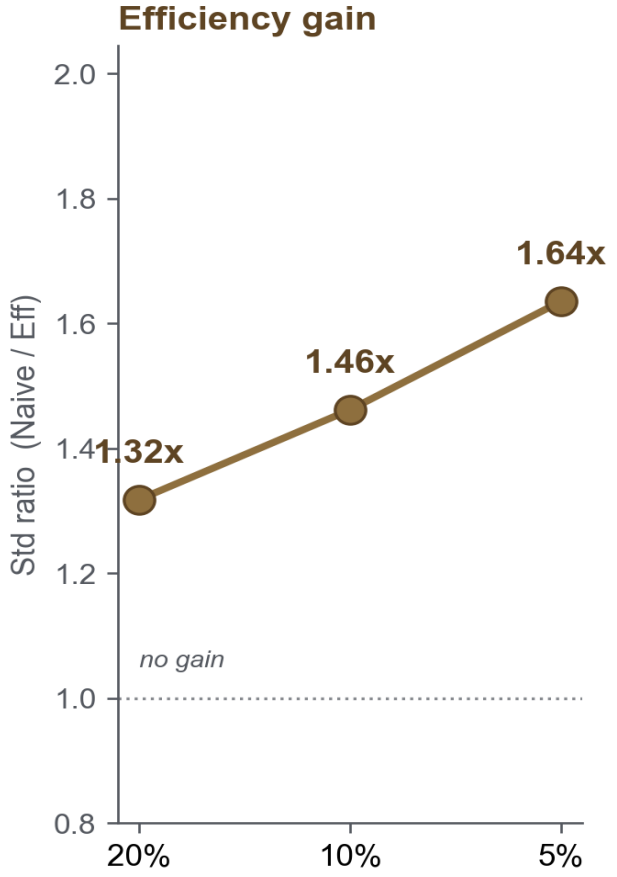
Target: $\psi_1 = T^*[Gemini, math]$

PURDUE UNIVERSITY

Sample Efficiency Advantage Grows with Scarcity

T*[Gemini, math] = 0.687 — Naive vs Efficient (500 trials, 20 / 10 / 5 %)



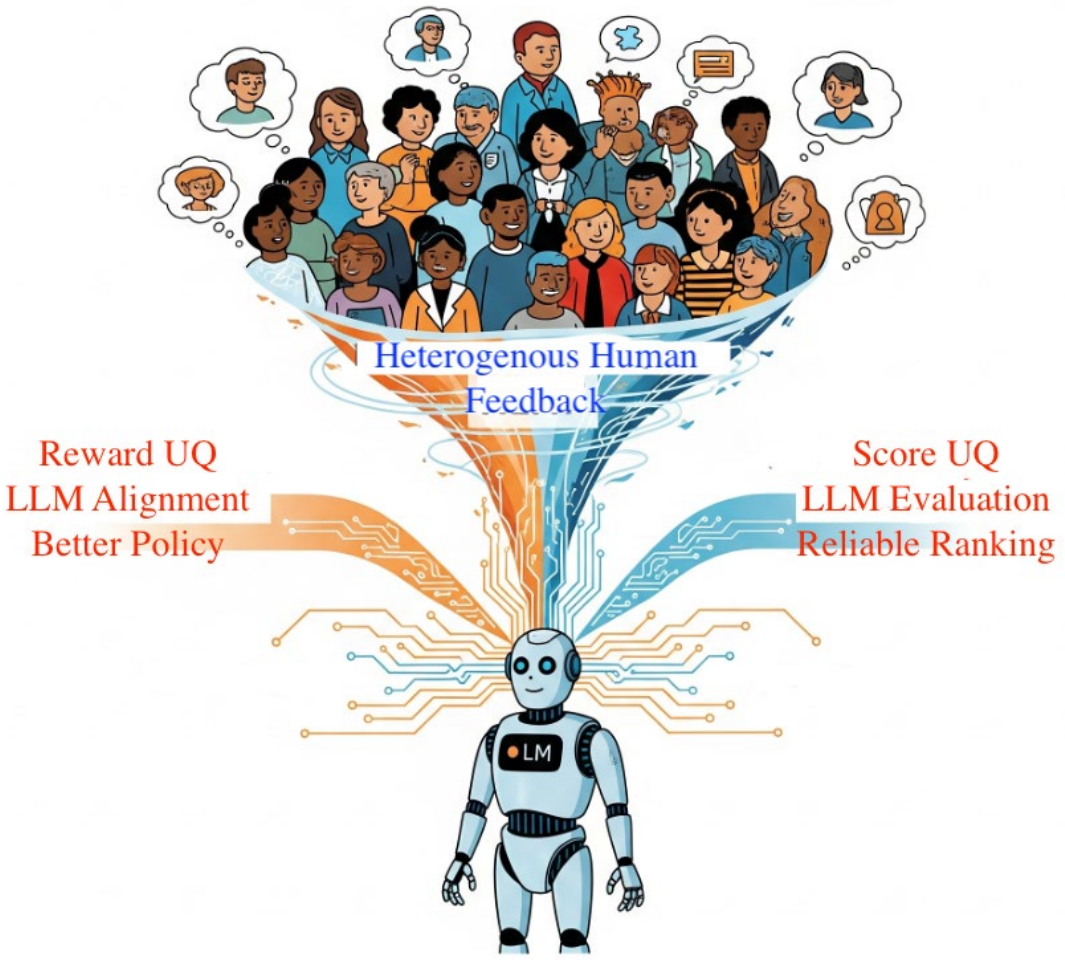


Coverage & rejection rate (H_0: $\psi \leq 0$)

Frac	Method	Cov	Rej
20%	Naive	0.998	81%
20%	Efficient	0.930	98%
10%	Naive	0.982	52%
10%	Efficient	0.924	80%
5%	Naive	0.972	32%
5%	Efficient	0.908	47%

- Efficiency gain grows steadily: 1.32x at 20% → 1.46x at 10% → 1.64x at 5%.
- At 10%, naive rejection rate drops to 52%; Efficient still rejects at 80%.
- Efficient coverage stays above 0.90 at all fractions.

Summary



Future direction:

Design-based active learning

Will Wei Sun
Purdue University
sun244@purdue.edu
Thanks NSF!