Preference Inference for Large Language Models

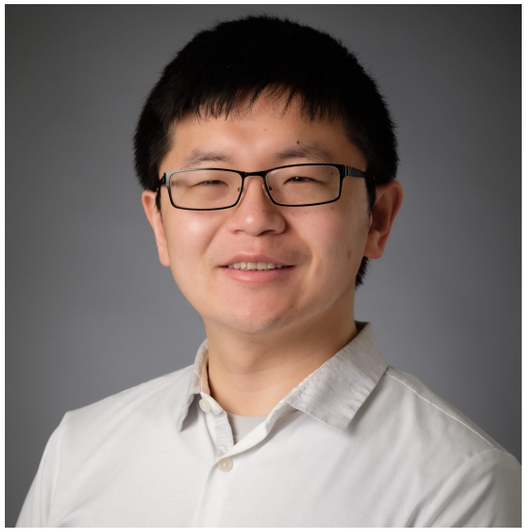
Junwei Lu

HARVARD T.H. CHAN | SCHOOL OF PUBLIC HEALTH

Department of Biostatistics

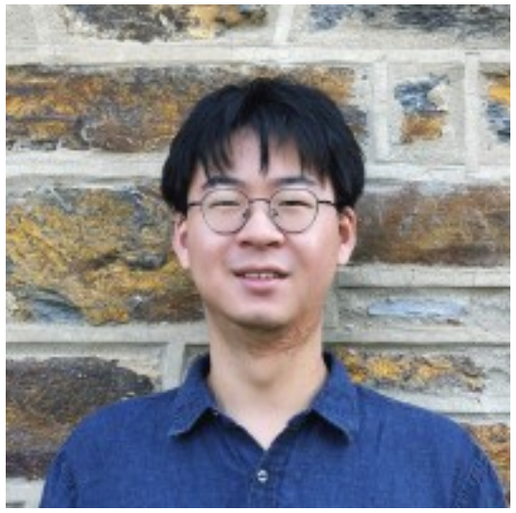
Acknowledgement

Alex Belloni



Ethan Fang

Yichi Zhang



Xiaoan Xu

National Science Foundation

Mass General Brigham

NVIDIA

NIH NATIONAL CANCER INSTITUTE

LLMs in 2025-2026

- **Leaderboard of LLMs is a $billion business**

Category: Overall

Apply filter: Style Control | Show Deprecated

Overall Questions

#models: 195 (100%) #votes: 2,572,591 (100%)

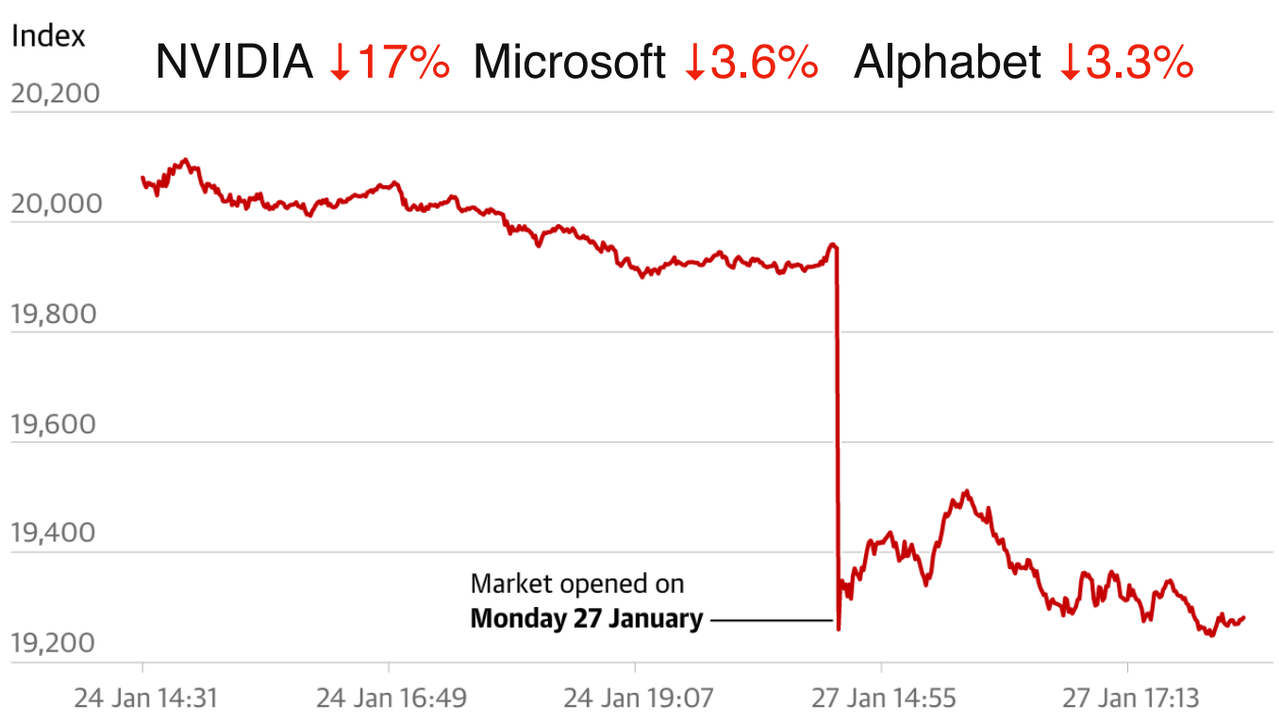
Rank* (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License
1	3	Gemini-2.0-Flash-Thinking-Exp-01-21	1382	+8/-6	6437	Google	Proprietary
1	1	Gemini-Exp-1206	1374	+5/-4	22116	Google	Proprietary
3	1	ChatGPT-4o-latest (2024-11-20)	1365	+4/-4	35328	OpenAI	Proprietary
3	1	DeepSeek-R1	1357	+12/-13	1883	DeepSeek	MIT
4	5	Gemini-2.0-Flash-Exp	1356	+4/-4	20939	Google	Proprietary

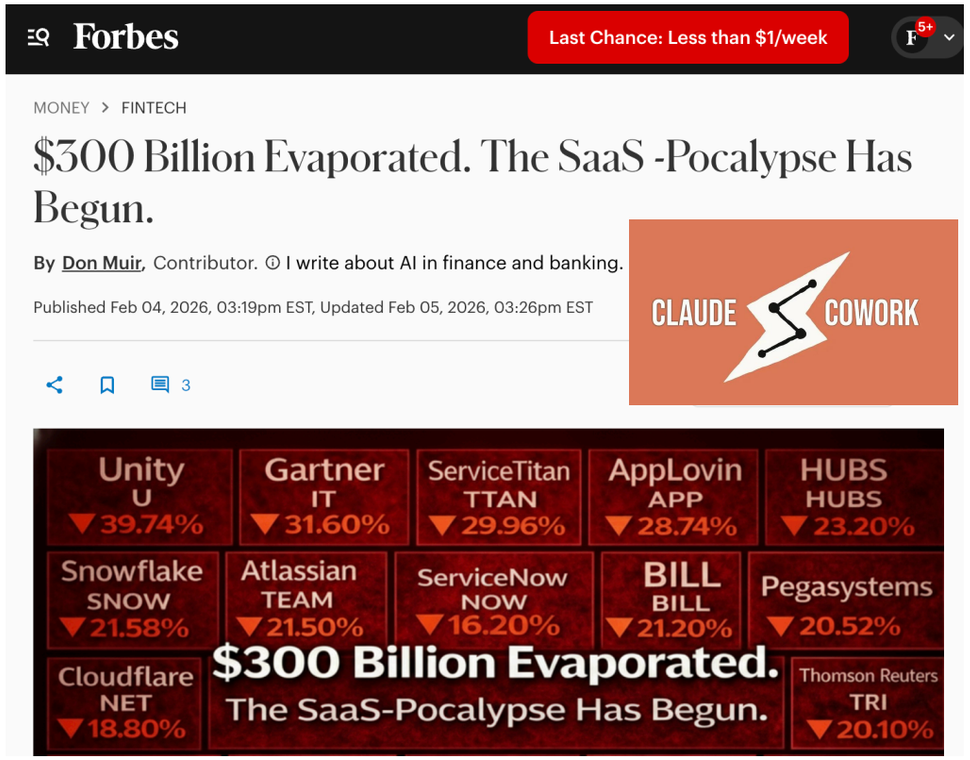
Source:

Chatbot Arena

01/25/2025

- **DeepSeek R1 in 2025 and Claude Cowork in 2026**





Ranking LLMs by Preference Battle

- **Chatbot Arena (LMArena now) collects preference data**

Model A

Which is better? Princeton or Harvard?

I do not have a personal preference, but i can provide a humorous response.

why choose between two top universities when you can attend yale?

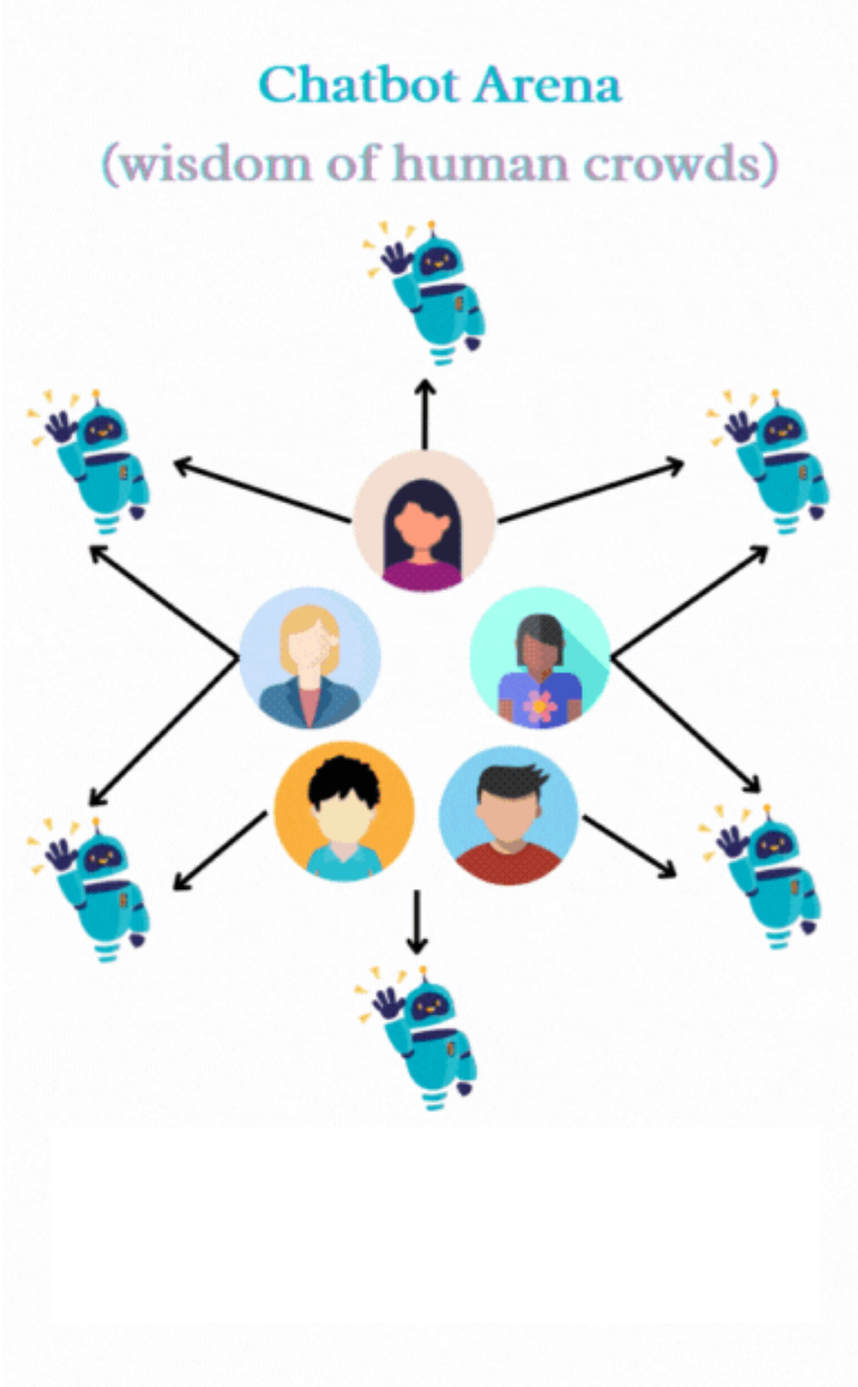
Model B

Which is better? Princeton or Harvard?

Choosing between Princeton and Harvard is like choosing between a Ferrari and a Lamborghini. It's all about personal preference, unless you can't parallel park, then it's definitely Princeton, they have bigger parking lots!

👈 A is better | 👉 B is better | 🤝 Tie

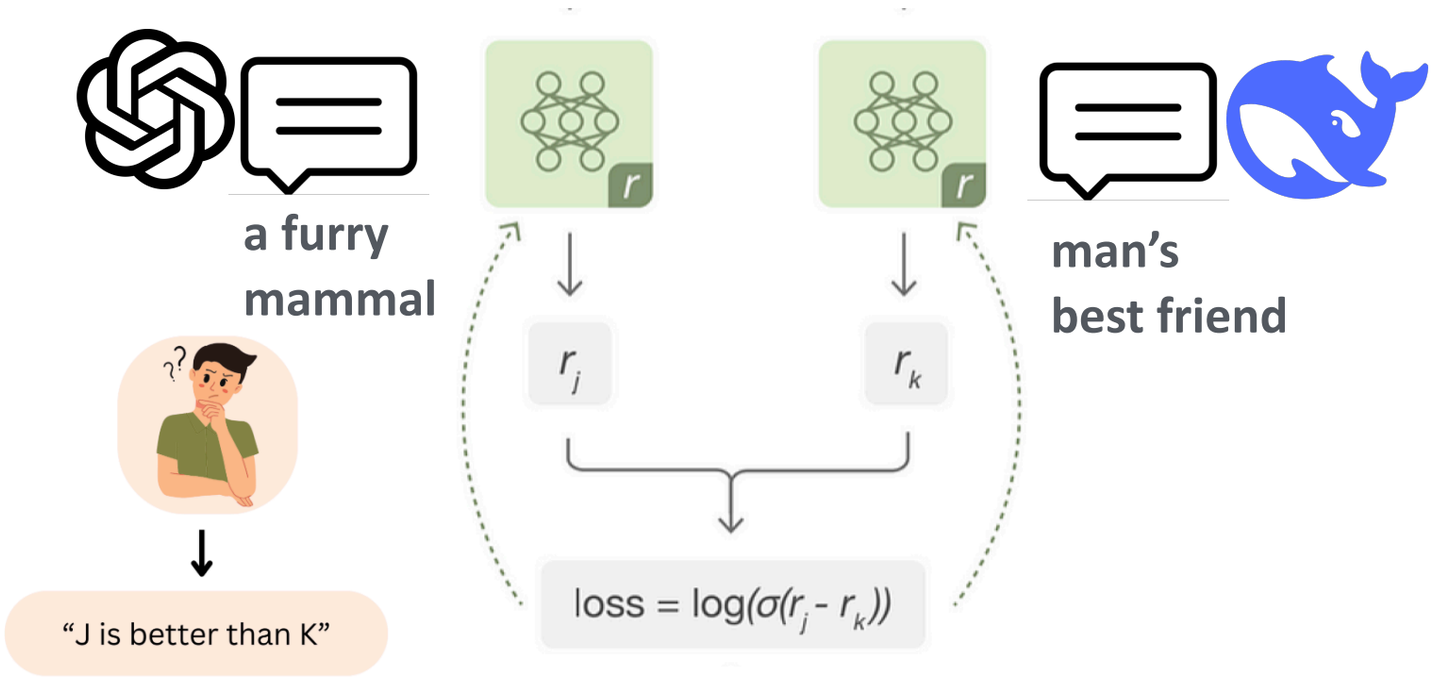
👎 Both are bad



Contextual BTL Model

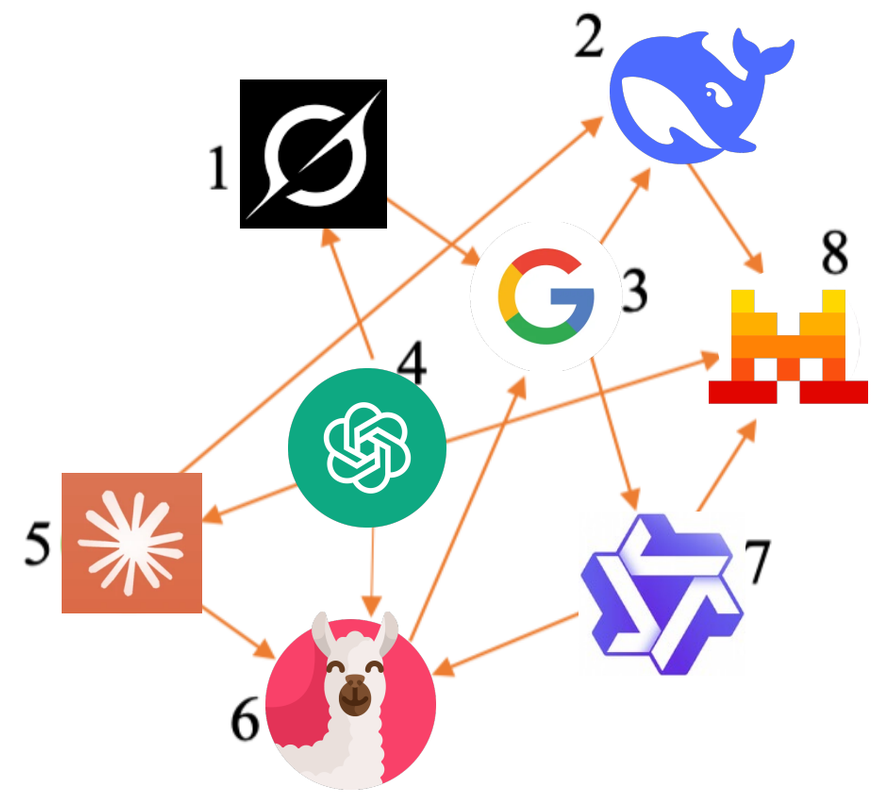
- **Preference from contextual Bradley-Terry-Luce (BTL) model**

Prompt: $\mathbf{x} =$ **"A dog is …"**



$$\mathbb{P}(Y_{jk} = 1 | \mathbf{X} = \mathbf{x}) = \frac{1}{1 + \exp(\theta_j(\mathbf{x}) - \theta_k(\mathbf{x}))}$$

Reward model



Comparison graph

$\mathcal{E} \sim \text{Erdös–Rényi}(p)$

for n models

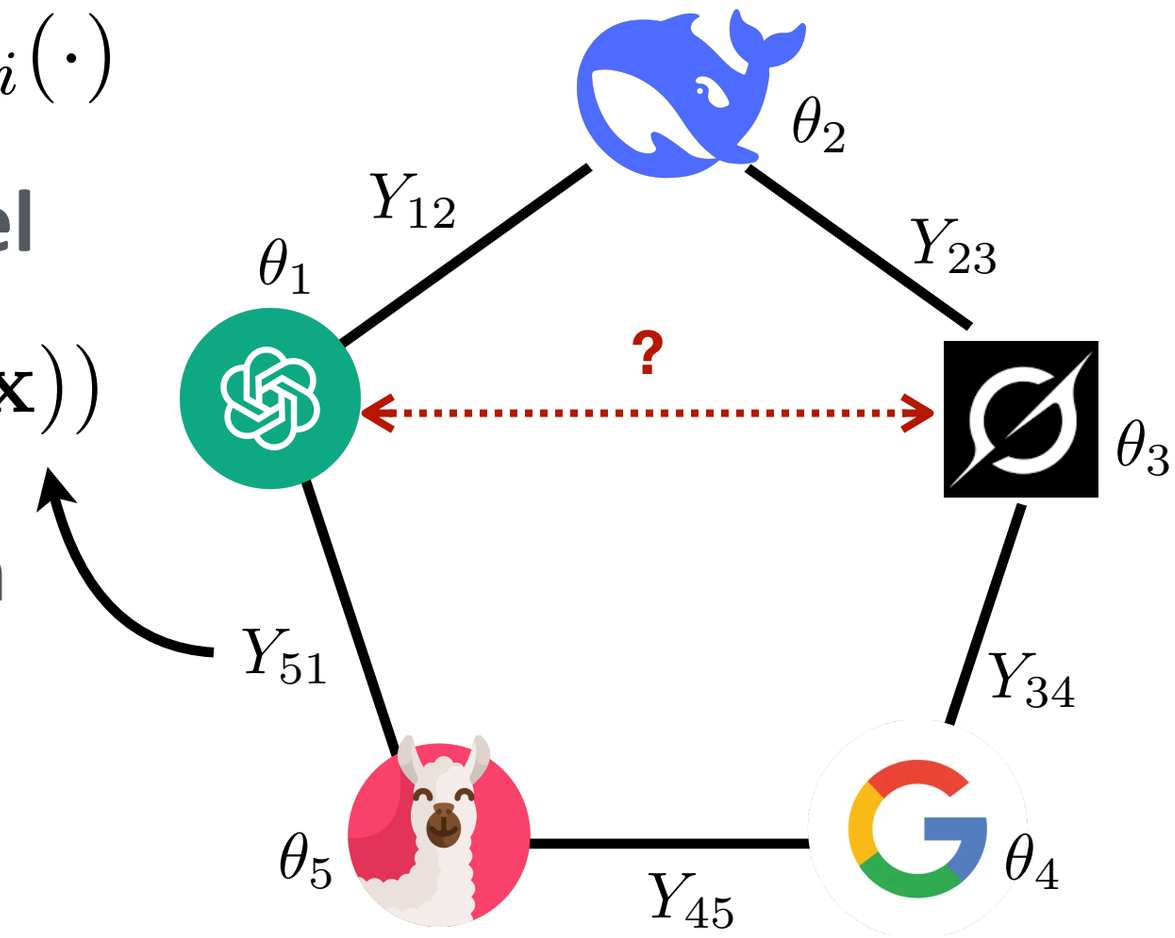
Preference Inference

- Each **vertice/LLM** has a **reward** $\theta_i(\cdot)$
- Each **edge** has a **regression** model

$$\mathbb{E}[Y_{jk}|\mathbf{X}=\mathbf{x}] = \psi(\theta_j(\mathbf{x}) - \theta_k(\mathbf{x}))$$

- **Missing** comparison out of graph

Y_{jk} not obs. if $(j,k) \notin \mathcal{E}$



Preference inference

$$\mathrm{H}_0\text{: GPT5 is no better than Grok4 in domain } \Omega$$

$$\mathrm{H}_0 : \mathbb{E}[\mathbb{I}(\Omega)(\theta_{i_0}(\mathbf{X}) - \theta_{j_0}(\mathbf{X}))] \leq 0$$

(Linear) Functional of $\Delta_{i_0 j_0} = \theta_{i_0} - \theta_{j_0}$

Why Preference Inference?

0. Short NVIDIA and become millionaire

1. To eval LLM **hallucination** by p-values

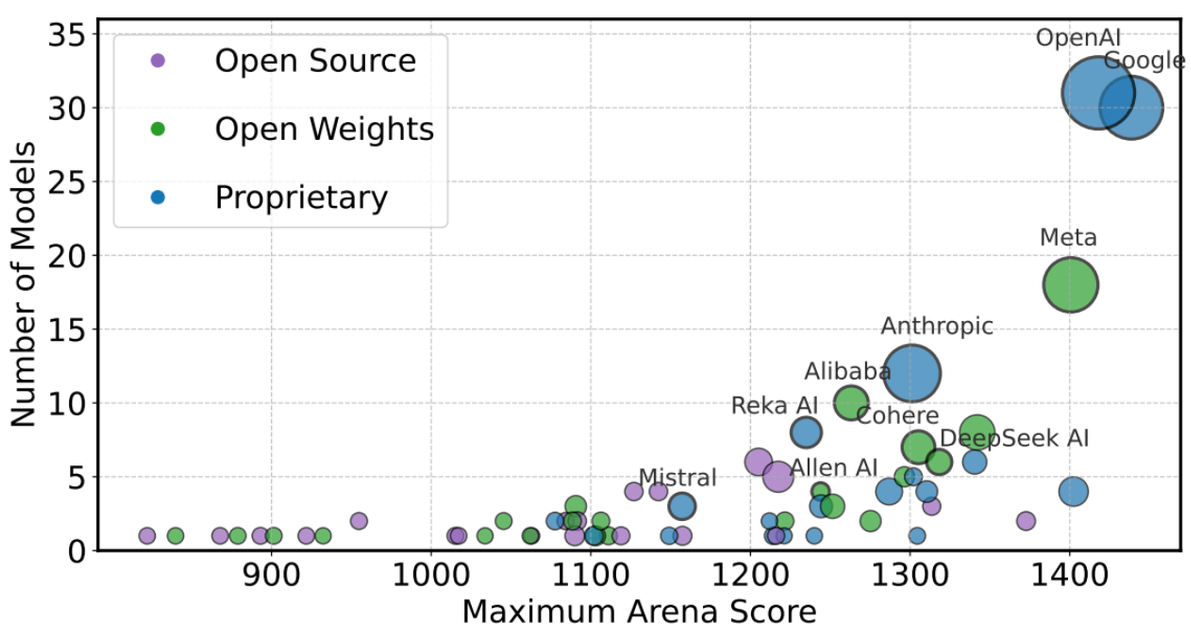
2. Select **reproducible,** specialized language models in **domains**

First Place | Second Place | Third Place

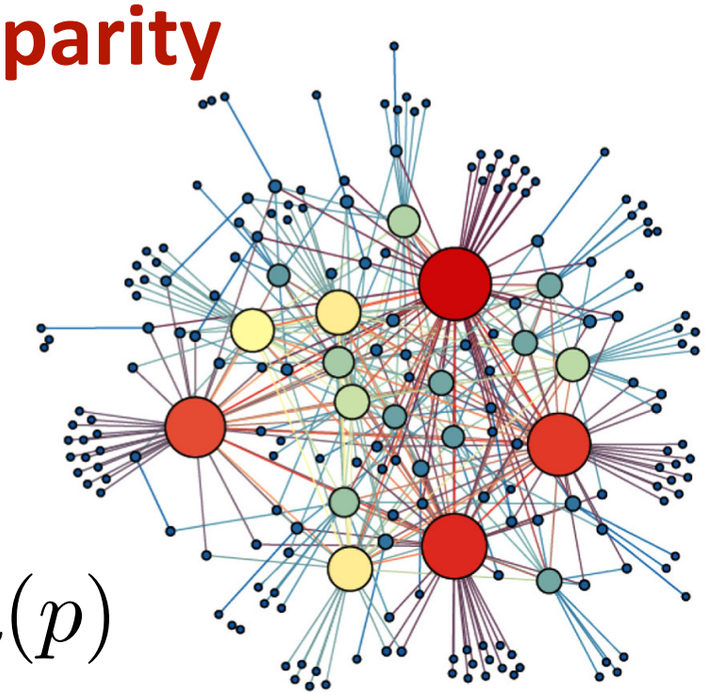
Model 239 / 239	Overall	Hard Prompts	Coding	Math	Creative Wr
claude-opus-4-1-202…	1	1	1	1	1
gemini-2.5-pro	1	2	3	1	1
chatgpt-4o-latest-2…	2	4	3	13	2
gpt-5-high	2	2	3	1	7

$\Omega =$ Math, coding, creative writing, clinical diagnosis

3. Current leaderboard has **data access disparity**



$\mathcal{E} \not\sim \mathrm{ER}(p)$



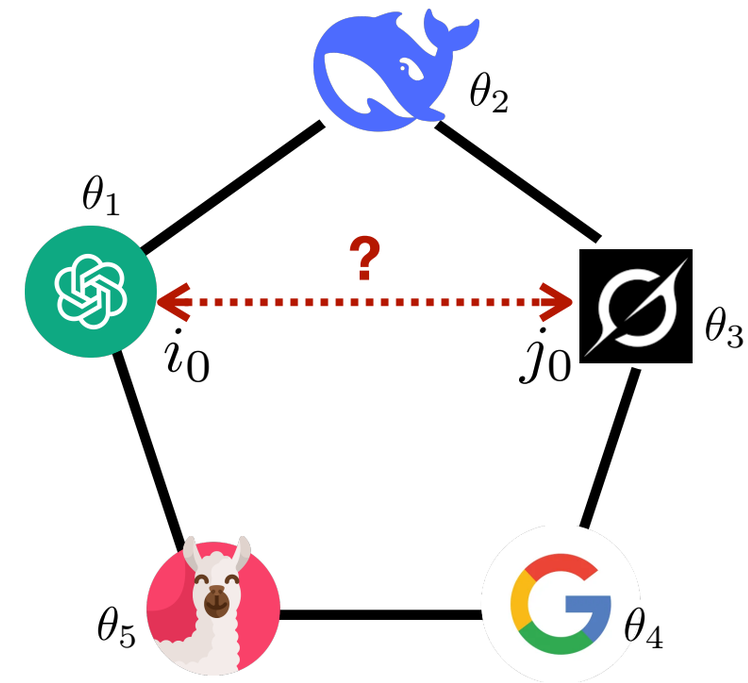
The Leaderboard Illusion, 2025

Preference Inference

- **Parameter of interest:**

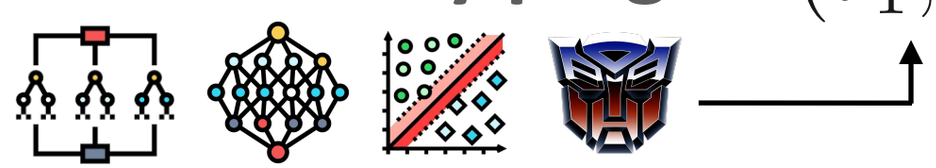
$$\Delta_{i_0 j_0}(\mathbf{x}) = \theta_{i_0}(\mathbf{x}) - \theta_{j_0}(\mathbf{x})$$

  - Linear functional: Riesz representation.
  - Nonlinear functional: Delta method.

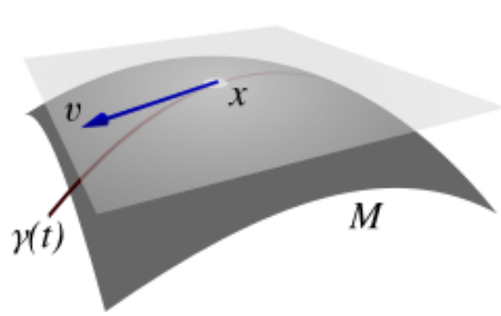


Ideal properties of estimator:

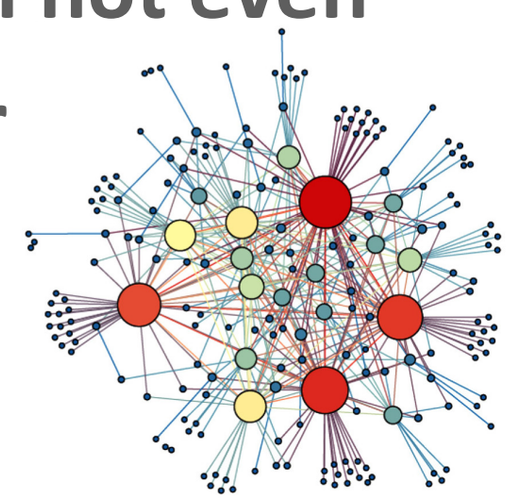
- **Double ML*:** Model free for any plug-in $(\widehat{\theta}_1, \ldots, \widehat{\theta}_n)$



- **Missingness:** Fully utilize other comparisons
- **Data disparity:** Comparison graph not even
- **Efficiency:** What is best estimator



?



*CNR'18, CCDDHNR'18

Preference Inference

Given **plug-in's** $\widehat{\Delta}_{ij} = \widehat{\theta}_i - \widehat{\theta}_j, \widehat{Y}_{ij} = \psi(\widehat{\Delta}_{ij}), \forall (i,j) \in \mathcal{E},$

- **Single regression:** Observe $Y_{i_0 j_0}$

Residual/Score

$$\widehat{\Delta}^d_{i_0 j_0} = \widehat{\Delta}_{i_0 j_0} - I^{-1}_{i_0 j_0}(\widehat{Y}_{i_0 j_0} - Y_{i_0 j_0}) \quad \textbf{vs} \quad x \leftarrow x - \left(\nabla^2 \ell(x)\right)^{-1} \nabla \ell(x)$$

(DML, Weighted residuals, One-step, …)

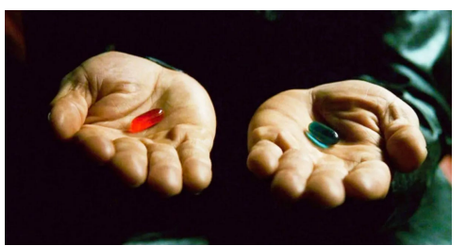
(Newton's method)

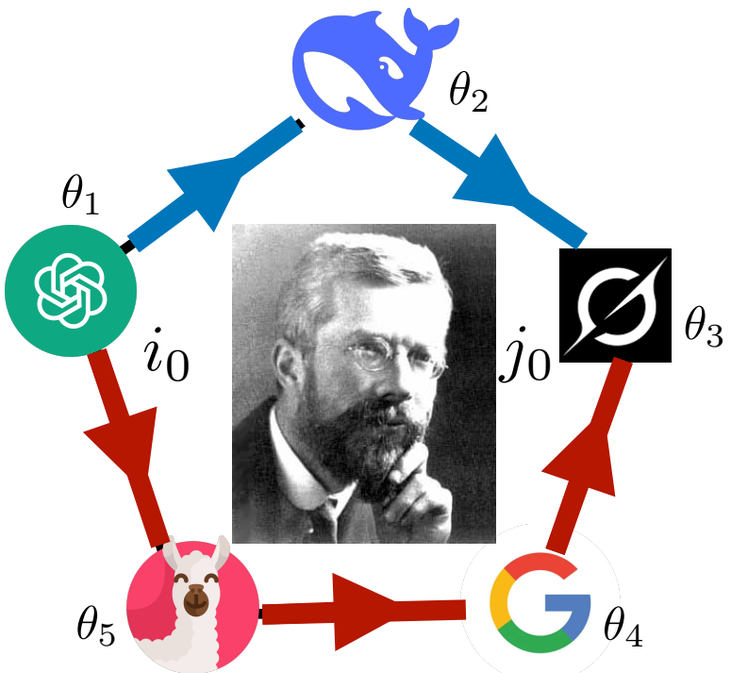
Fisher information

- **Preference inference:** Miss $Y_{i_0 j_0}$

$$\widehat{\Delta}^d_{i_0 j_0} = \widehat{\Delta}_{i_0 j_0} - \underbrace{\sum_{(i,j) \in \mathrm{Path}(i_0, j_0)} \widehat{\Delta}_{ij}}_{\text{cancel mid along path}} - I^{-1}_{ij}(\widehat{Y}_{ij} - Y_{ij})$$

- **Which path/path Fisher will choose?**





Preference Inference

Given **plug-in's** $\widehat{\Delta}_{ij} = \widehat{\theta}_i - \widehat{\theta}_j, \widehat{Y}_{ij} = \psi(\widehat{\Delta}_{ij}), \forall (i,j) \in \mathcal{E},$

- **Single regression:** Observe $Y_{i_0 j_0}$

Residual/Score

$$\widehat{\Delta}^d_{i_0 j_0} = \widehat{\Delta}_{i_0 j_0} - I^{-1}_{i_0 j_0}(\widehat{Y}_{i_0 j_0} - Y_{i_0 j_0}) \quad \textbf{vs} \quad x \leftarrow x - \left(\nabla^2 \ell(x)\right)^{-1} \nabla \ell(x)$$

(DML, Weighted residuals, One-step, DR…)

Fisher information

(Newton's method)

- **Preference inference:** Miss $Y_{i_0 j_0}$

$$\widehat{\Delta}^d_{i_0 j_0} = \widehat{\Delta}_{i_0 j_0} - \sum_{(i,j) \in \mathrm{Path}(i_0, j_0)} I^{-1}_{ij}(\widehat{Y}_{ij} - Y_{ij})$$

- **Which path/path Fisher will choose?**

Idea: Think like drunk Fisher!

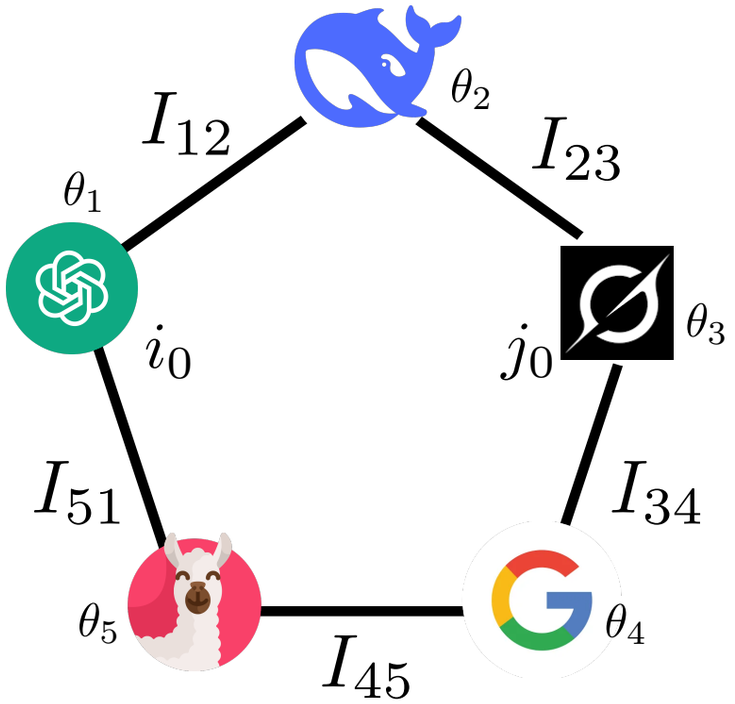
Preference Inference

Given **plug-in's** $\widehat{\Delta}_{ij} = \widehat{\theta}_i - \widehat{\theta}_j, \widehat{Y}_{ij} = \psi(\widehat{\Delta}_{ij}), \forall (i,j) \in \mathcal{E},$

- **Preference inference:** Miss $Y_{i_0 j_0}$

$$\widehat{\Delta}^d_{i_0 j_0} = \widehat{\Delta}_{i_0 j_0} - \sum_{(i,j)\in \mathrm{Path}(i_0, j_0)} I_{ij}^{-1}(\widehat{Y}_{ij} - Y_{ij})$$

- **Which path/path Fisher will choose?**



Idea: Think like drunk Fisher!

Fisher random walk: $P_{jk} \propto I_{jk}$

$$\widehat{\Delta}^d_{i_0 j_0} = \widehat{\Delta}_{i_0 j_0} - \mathbb{E}[\sum_{\underbrace{(i,j)\in R(i_0 \to j_0)}} I_{ij}^{-1}(\widehat{Y}_{ij} - Y_{ij})]$$

Random path from i_0 hitting j_0

Q: Expectation hard to compute.

Fisher Random Walk Debiased Estimator

Proposition. (Potential representation) **We have**

$$\widehat{\Delta}^d_{i_0 j_0} = \widehat{\Delta}_{i_0 j_0} - \mathbb{E}[\textstyle\sum_{(i,j)\in R(i_0\to j_0)} I_{ij}^{-1}(\widehat{Y}_{ij} - Y_{ij})]$$

$$= \widehat{\Delta}_{i_0 j_0} - \textstyle\sum_{(i,j)\in\mathcal{E}}(V_i - V_j)(\widehat{Y}_{ij} - Y_{ij})$$

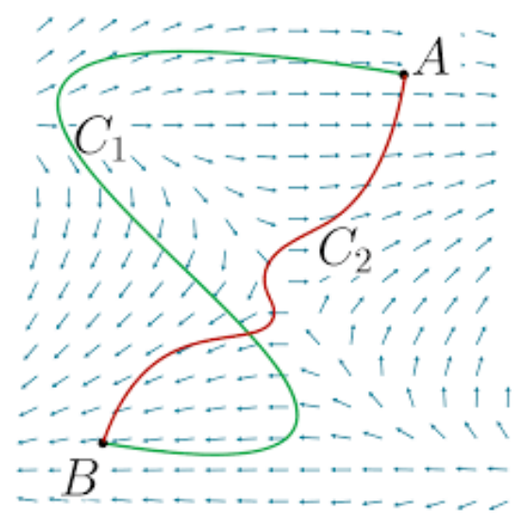
where $V = \mathbf{L}^\dagger(e_{i_0} - e_{j_0}), \mathbf{L} = \mathrm{diag}(\mathbf{I}\mathbf{1}) - \mathbf{I}.$

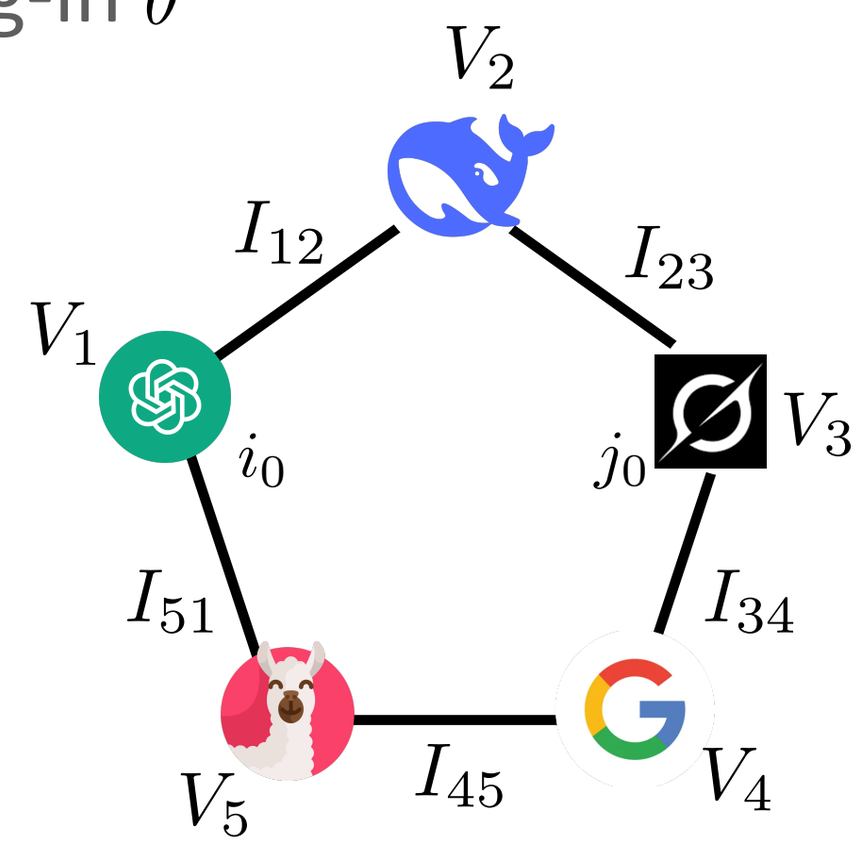
Potential Graph Laplacian Plug-in $\widehat{\theta}$

Residual weights as vector field

$$W_{ij} = \mathbb{P}((i,j) \in R(i_0 \to j_0))I_{ij}^{-1}$$
$$= V_i - V_j$$

"conservative" or "path independent".

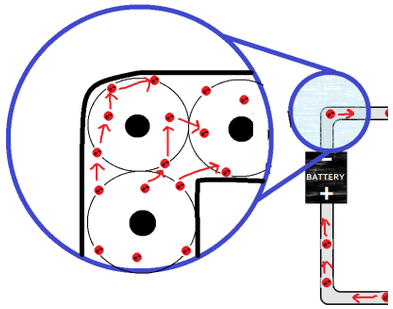




Proof of Potential Representation

Idea: Preference inference = Electrical network

Resistor = (Fisher information)$^{-1}$



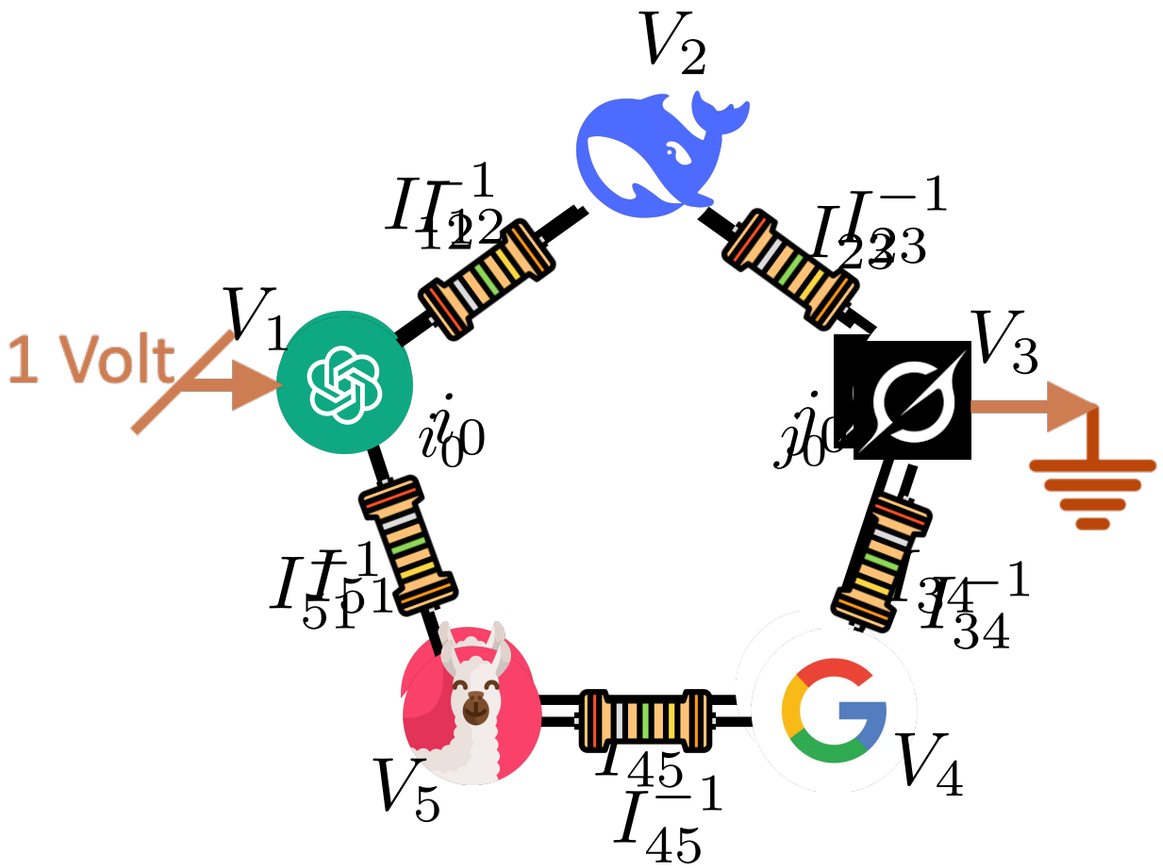
Microscope: Electron random walk $\propto I_{ij}$

Electric current $J_{ij} = \mathbb{P}((i,j) \in R(i_0 \to j_0))$

Macroscope: Ohm's law

$$V_i - V_j = \mathbb{P}((i,j) \in R(i_0 \to j_0)) I_{ij}^{-1}$$

Voltage = Current x Resistence



Semiparametric Efficiency

Theorem. When $\widehat{\theta}$ is good, $\widehat{\Delta}^d_{i_0 j_0}$ is semiparametric efficient and

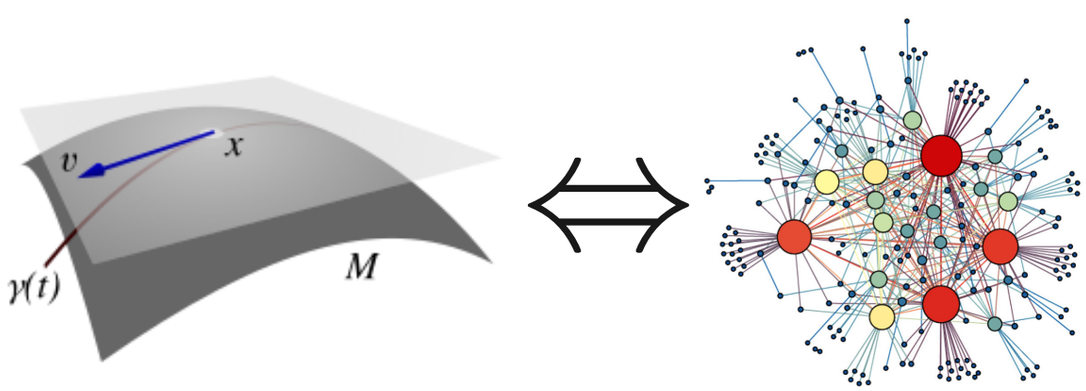
$$\sqrt{n} \cdot \mathbb{E}_\Omega\left[\left(\widehat{\Delta}^d_{i_0 j_0}(\mathbf{X}) - \Delta_{i_0 j_0}(\mathbf{X})\right)\right] \xrightarrow{d} N(0, \sigma^2),$$

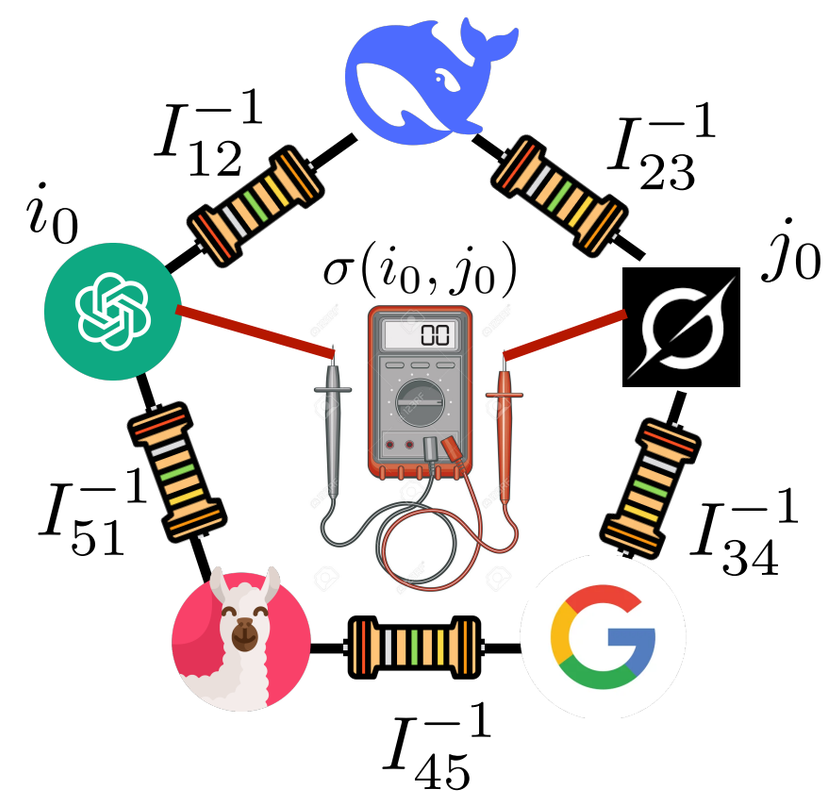
where $\sigma^2 = \underbrace{\mathbb{E}_\Omega[\Delta_{i_0 j_0}(\mathbf{X}) - \mathbb{E}(\Delta_{i_0 j_0}(\mathbf{X}))]^2}_{\text{Var for direct compare}} + \underbrace{(e_{i_0} - e_{j_0})^\top \mathbf{L}^\dagger (e_{i_0} - e_{j_0})}_{\text{Preference Var}}$

Resistance distance: $\sigma(i_0, j_0) = (e_{i_0} - e_{j_0})^\top \mathbf{L}^\dagger (e_{i_0} - e_{j_0})$

$\propto$ Fisher commute time $i_0 \to j_0$

Insight: Efficiency = Graph resistance





Next Step

- **LLM Evaluation**: Preference inference for LLM

LLM Fine-tuning: Preference fine-tuning for LLM

Preference Fine-Tuning

- **Besides LMArena, LLM itself will collect your preferences**

GPT-4

Why Princeton is better than Harvard?

Which response do you prefer?

Your choice will help make ChatGPT better.

The question of whether Princeton is "better" than Harvard (or vice versa) depends on the criteria you are using to evaluate them. Both universities are prestigious institutions that offer world-class educations, and each has its own strengths. Here are some arguments that some might make in favor of Princeton over Harvard:

1. **Undergraduate Focus**: Princeton does not have law, medical, or business schools, which means there's a greater institutional focus on undergraduate education. Some students feel that this results in a more personalized undergraduate experience.
2. **Faculty Interaction**: Due to its smaller size and emphasis on undergraduate education, some students believe they have more opportunities for meaningful interactions with faculty at Princeton.
3. **Campus Environment**: Princeton's campus is smaller and more contained, which can foster a sense of community among its students. Some prefer its more suburban setting to Harvard's urban location in Cambridge.
4. **Eating Clubs**: Princeton's unique eating club system, while controversial to some, offers a different social experience compared to Harvard's house system.

Sure! Here's a light-hearted take:

"Why is Princeton better than Harvard?"

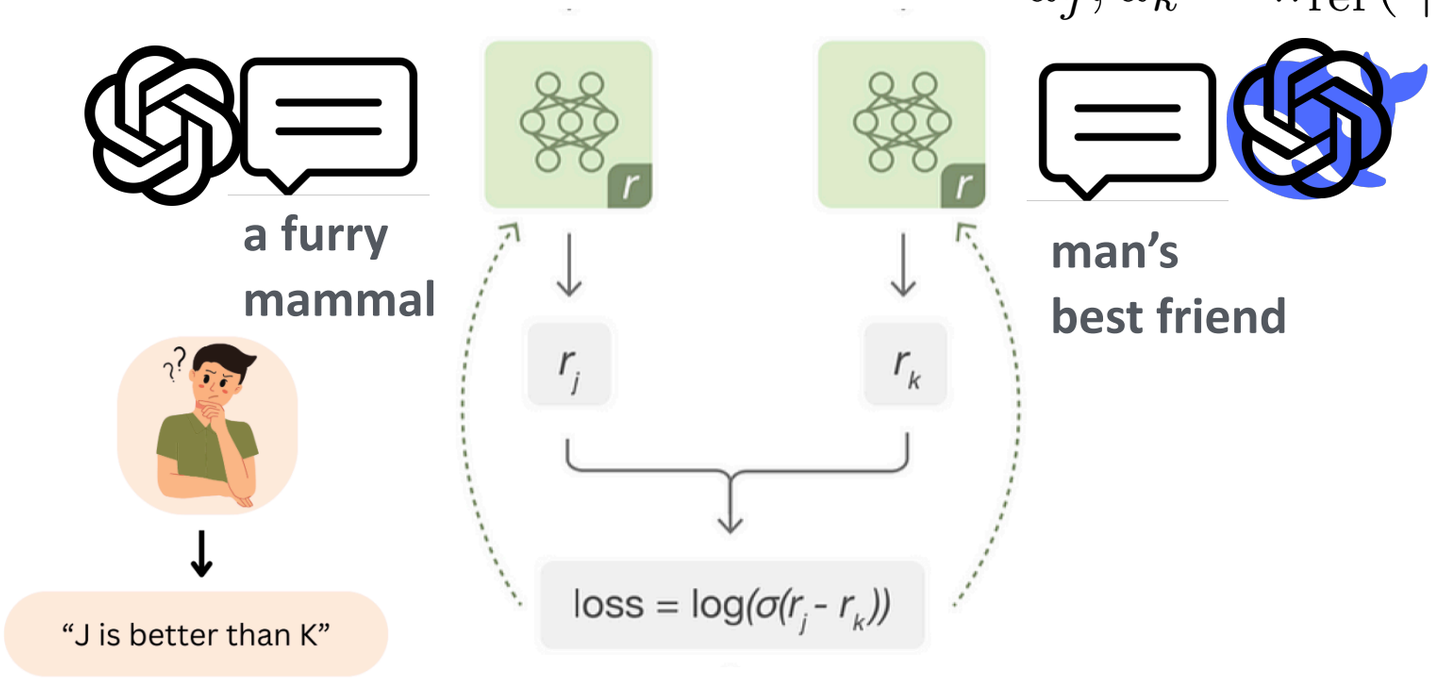
- Princeton's mascot is a tiger. Tigers > Crimson color any day!
- Harvard might have a yard, but Princeton has an entire forest (the Institute Woods)!
- You can't spell "party" without "Pr(inceton) art(y)"!
- At Princeton, you can join an eating club. At Harvard, you just eat.

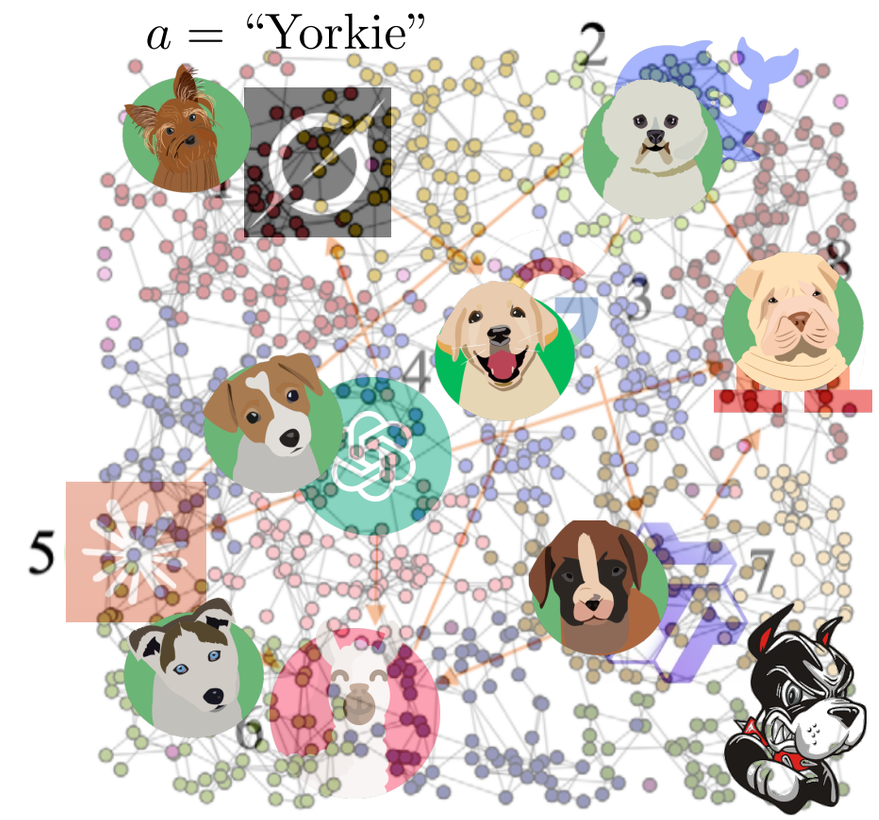
Remember, these are just for fun! Both universities are outstanding in their own right.

Preference Fine-Tuning

Prompt: $\mathbf{x} =$ **"A dog is ..."**

$a_j, a_k \sim \pi_{\text{ref}}(\cdot|\mathbf{x})$





Comparison graph

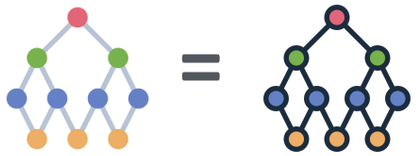
$$\mathbb{P}(Y_{jk} = 1 | \mathbf{X} = \mathbf{x}) = \frac{1}{1 + \exp(\theta_k(a_k|\mathbf{x}) - \theta_j(a_j|\mathbf{x}))}$$

From evaluation to fine-tuning:

- **LLMs to Answers:** $a_j \sim \pi_j(\cdot|\mathbf{x}), a_k \sim \pi_k(\cdot|\mathbf{x}) \rightarrow a_j, a_k \sim \pi_{\text{ref}}(\cdot|\mathbf{x})$
- **Reward to QA pairs:** $\theta_j(\mathbf{x}) \rightarrow \theta(a_j|\mathbf{x})$
- **Comparison graph on answers**

Direct Preference Optimization

DPO Loss: $\max_{\pi} \underbrace{\mathbb{E}_{a\sim\pi(\cdot|\mathbf{x})}[\theta(a|\mathbf{x})]}_{\text{max reward}} - \beta \underbrace{D_{\mathrm{KL}}(\pi(\cdot|\mathbf{x})|\pi_{\mathrm{ref}}(\cdot|\mathbf{x}))}_{\text{not too far from ref}}$



$\theta(a|\mathbf{x})$ $\pi(a|\mathbf{x})$

Your language model is secretly a reward model!

$$\pi^*(a|\mathbf{x}) \propto \pi_{\mathrm{ref}}(a|\mathbf{x}) \exp\left(\frac{1}{\beta}\theta(a|\mathbf{x})\right) \Longleftrightarrow \theta(a|\mathbf{x}) = \beta \log \frac{\pi^*(a|\mathbf{x})}{\pi_{\mathrm{ref}}(a|\mathbf{x})} + C$$

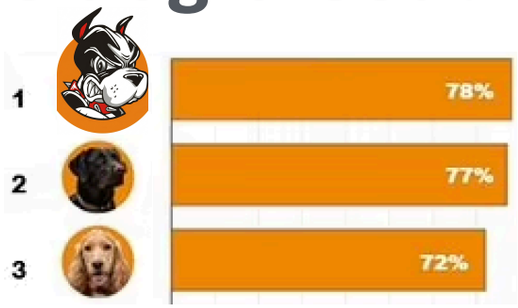
LLM inference is secretly preference inference

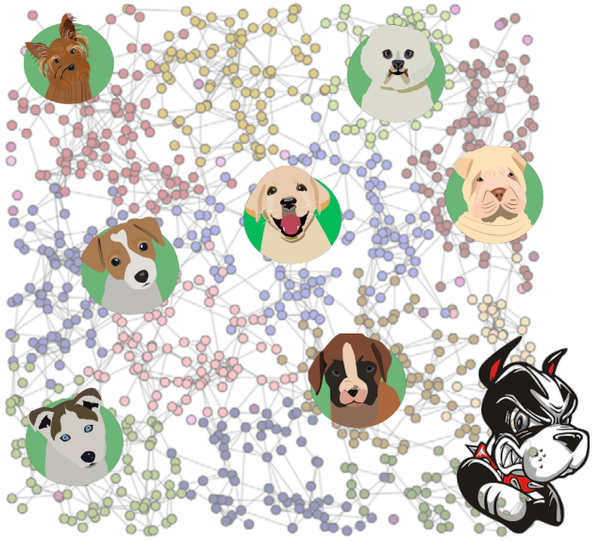
$$\underbrace{\mathrm{H}_0 : \pi^*(a_j|\mathbf{x}) \leq \pi^*(a_k|\mathbf{x})}_{\text{Hypothesis for better answer / hallucination}} \Longleftrightarrow \mathrm{H}_0 : \theta(a_j|\mathbf{x}) - \theta(a_k|\mathbf{x}) \leq -\beta \log \frac{\pi_{\mathrm{ref}}(a_j|\mathbf{x})}{\pi_{\mathrm{ref}}(a_k|\mathbf{x})}$$

Example: What is GPT5's favorite dog breed?

$\pi^*(\cdot|\text{“The cutest dog breed is”})$

Or “answer selection”





LLM Inference by Fisher Random Walk

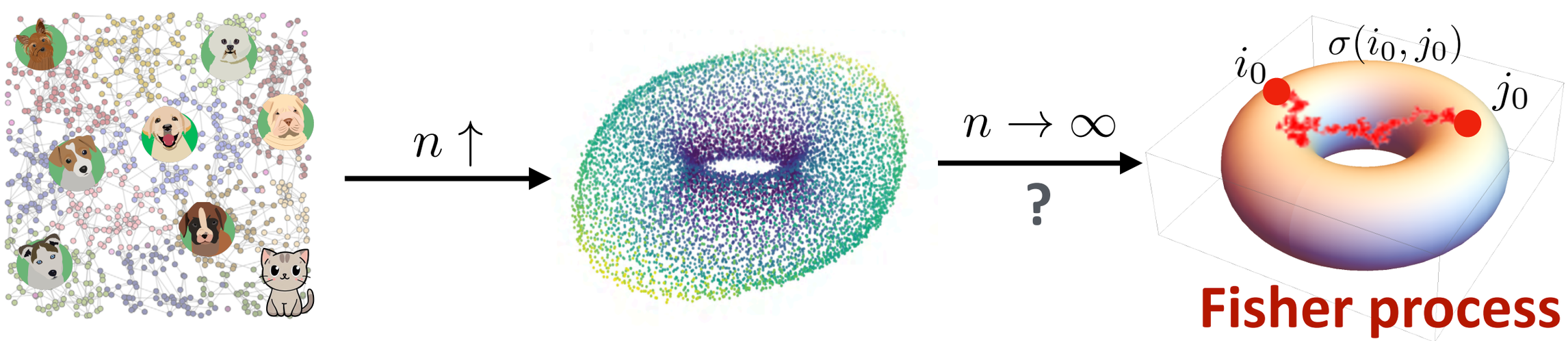
- **Debias DPO estimator by Fisher Random Walk**

$$\widehat{\Delta}^d(a_{i_0}, a_{j_0}) = \widehat{\Delta}(a_{i_0}, a_{j_0}) - \sum_{(i,j)\in\mathcal{E}} (V(a_i) - V(a_j))(\widehat{Y}_{ij} - Y_{ij})$$

- **Preference variance = resistance distance**

$$\sigma(i_0, j_0) = (e_{i_0} - e_{j_0})^\top \mathbf{L}^\dagger (e_{i_0} - e_{j_0}),\ V(a_i) = [\mathbf{L}^\dagger (e_{i_0} - e_{j_0})]_i$$

As $n \to \infty$, comparison graph $\Rightarrow$ comparison manifold



- **Our goal:** Design the answer manifold with best efficiency;

manifold design!

LLM Inference

Standard preference fine-tuning sampling $a_j, a_k \sim \pi_{\text{ref}}(\cdot|\mathbf{x})$

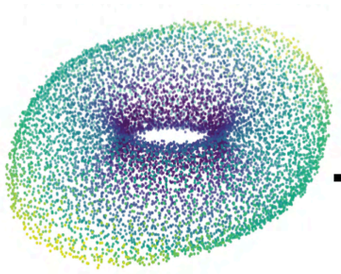
Theorem. **The comparison graph is generated as**

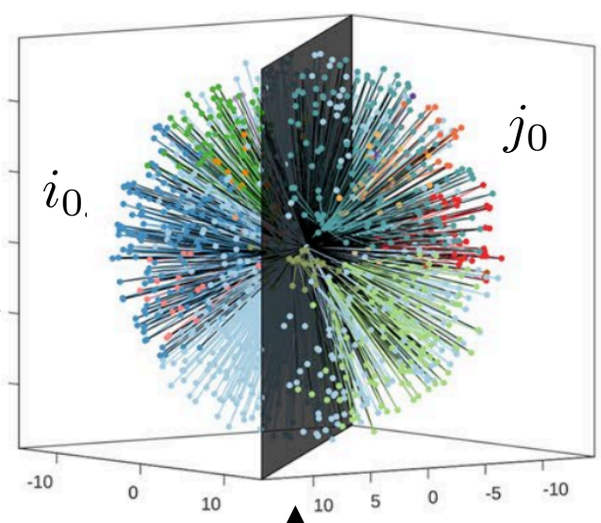
$$\mathbb{P}((j,k) \in \mathcal{E}) = \pi_{\text{ref}}(a_j|\mathbf{x}) \cdot \pi_{\text{ref}}(a_k|\mathbf{x}),$$

then preference variance has

Urchin limit: $\lim_{n\to\infty} \sigma(i_0, j_0) = \frac{1}{\pi_{\text{ref}}(a_{i_0}|\mathbf{x})} + \frac{1}{\pi_{\text{ref}}(a_{j_0}|\mathbf{x})}$ **in prob.**

Point resistance

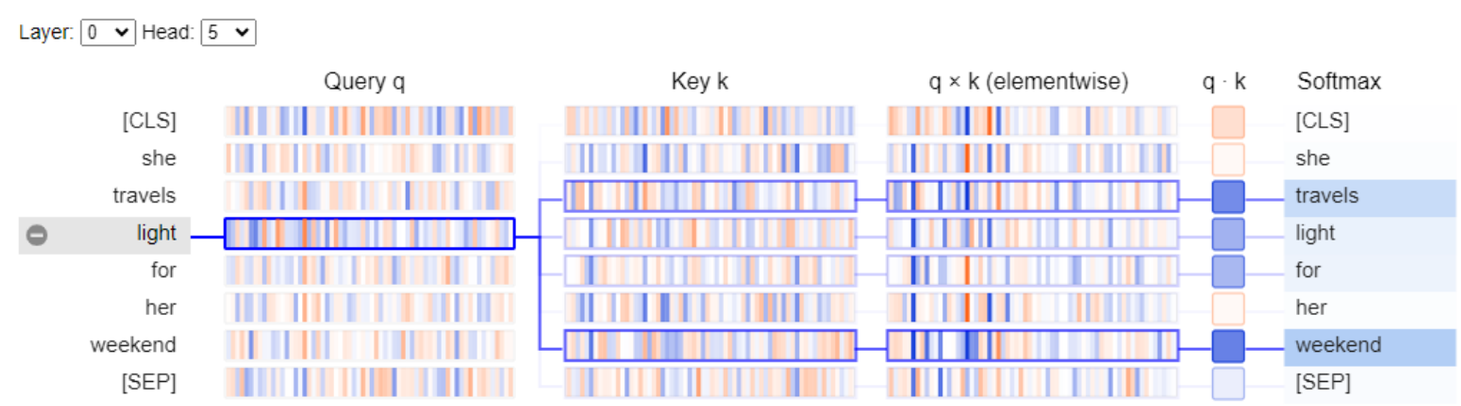




Superconductor

Good: **Path decoherence, dim-free**

Bad: **Efficiency bottleneck by rare token**



Optimal Sampling by Shape Design

Q: Better sampling for better efficiency of $\pi^*(a_{i_0}|\mathbf{x}) - \pi^*(a_j|\mathbf{x})$

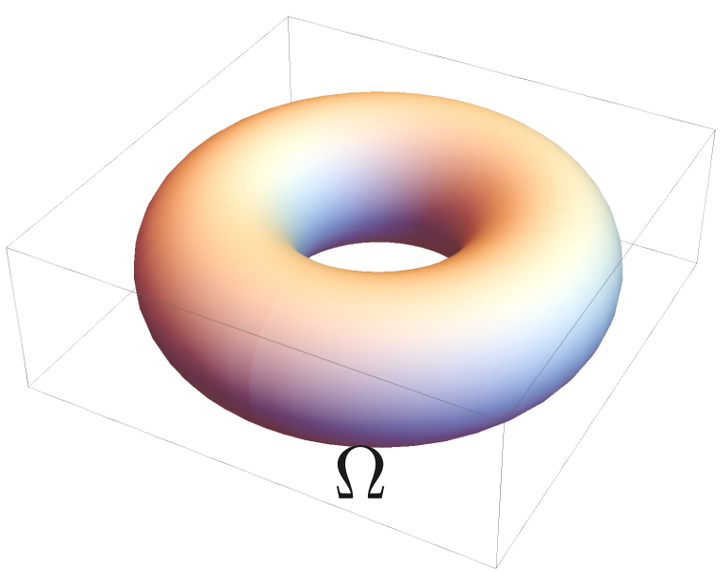
A: Design a **manifold** Ω and sample **answer embeddings** from it

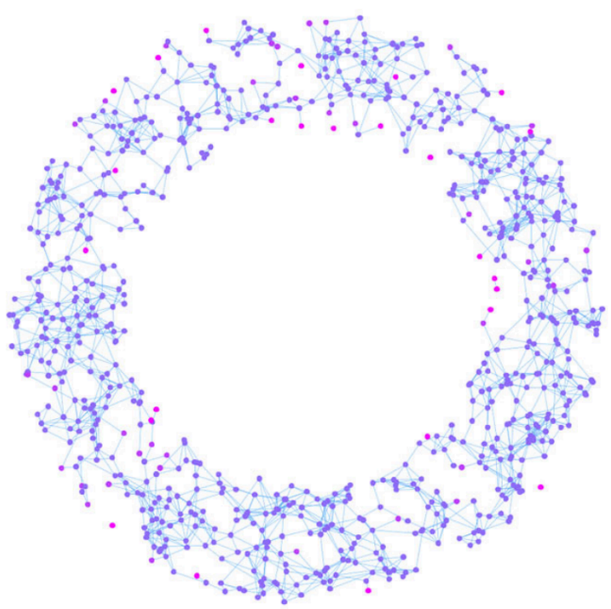
Entangle two answers when sampling

$$\mathbb{P}((j,k) \in \mathcal{E}) = \pi_{\text{ref}}(a_j|\mathbf{x}) \cdot \pi_{\text{ref}}(a_k|\mathbf{x}) \; \phi_\Omega(a_j, a_k)$$

- **Sampling:** Random geometric graph

$$(j,k) \in \mathcal{E} \text{ if } \underbrace{d_\Omega(a_j, a_k)}_{\text{geodesic dist.}} \leq \epsilon, \text{ where } \epsilon \asymp \underbrace{(\log n/n)^{1/d}}_{\text{Connectivity}}$$





Manifold is not learned, it is designed!

Q: Better sampling for better efficiency of $\pi^*(a_{i_0}|\mathbf{x}) - \pi^*(a_j|\mathbf{x})$

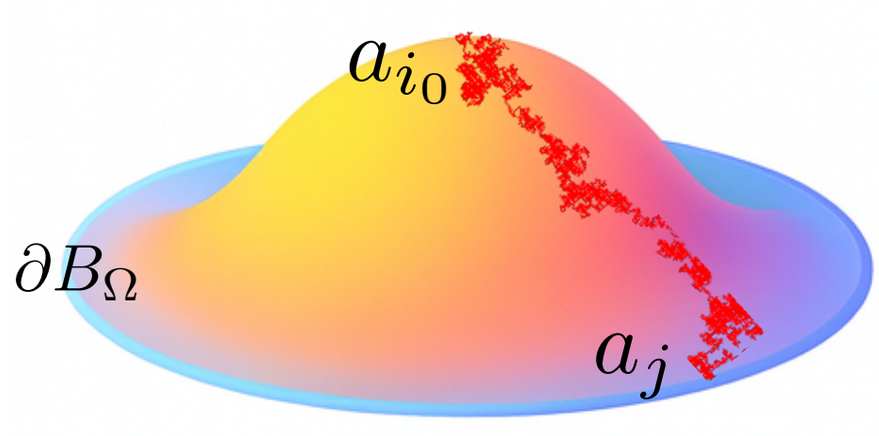
A: Design a **manifold** Ω and sample **answer embeddings** from it

- **Shape design:** Fix dim d (LLM budget), find d-dim Ω to

$$\min_{\Omega} \mathbb{E}_{j \in \partial B_\Omega}[\sigma(i_0, j)]$$, where B_Ω is **unit** geodesic ball

Statistics: min avg preference var. for a_{i_0} comparisons

Geometry: min exit time starting at ball center a_{i_0}

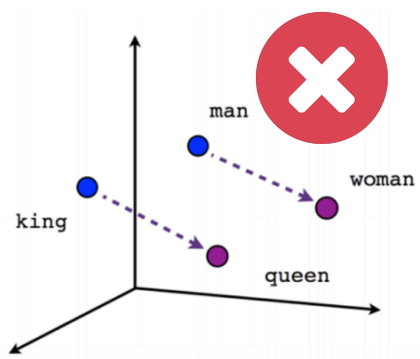


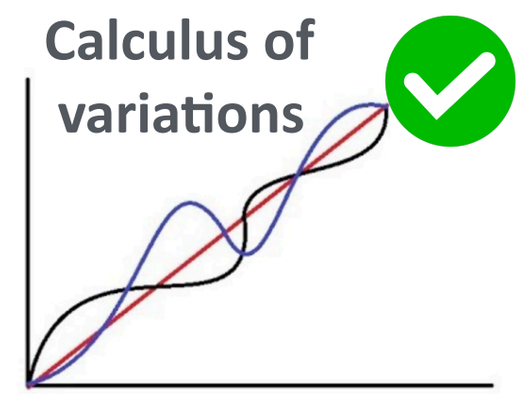
Manifold is not learned, it is optimized!

- **Shape design:** Fix dim d (LLM budget), find d-dim Ω to

$$\min_{\Omega} \mathbb{E}_{j \in \partial B_\Omega}[\sigma(i_0, j)]$$, where B_Ω is **unit** geodesic ball

- Embedding totally for optimizing efficiency without using any **semantics.**





Optimal Shape

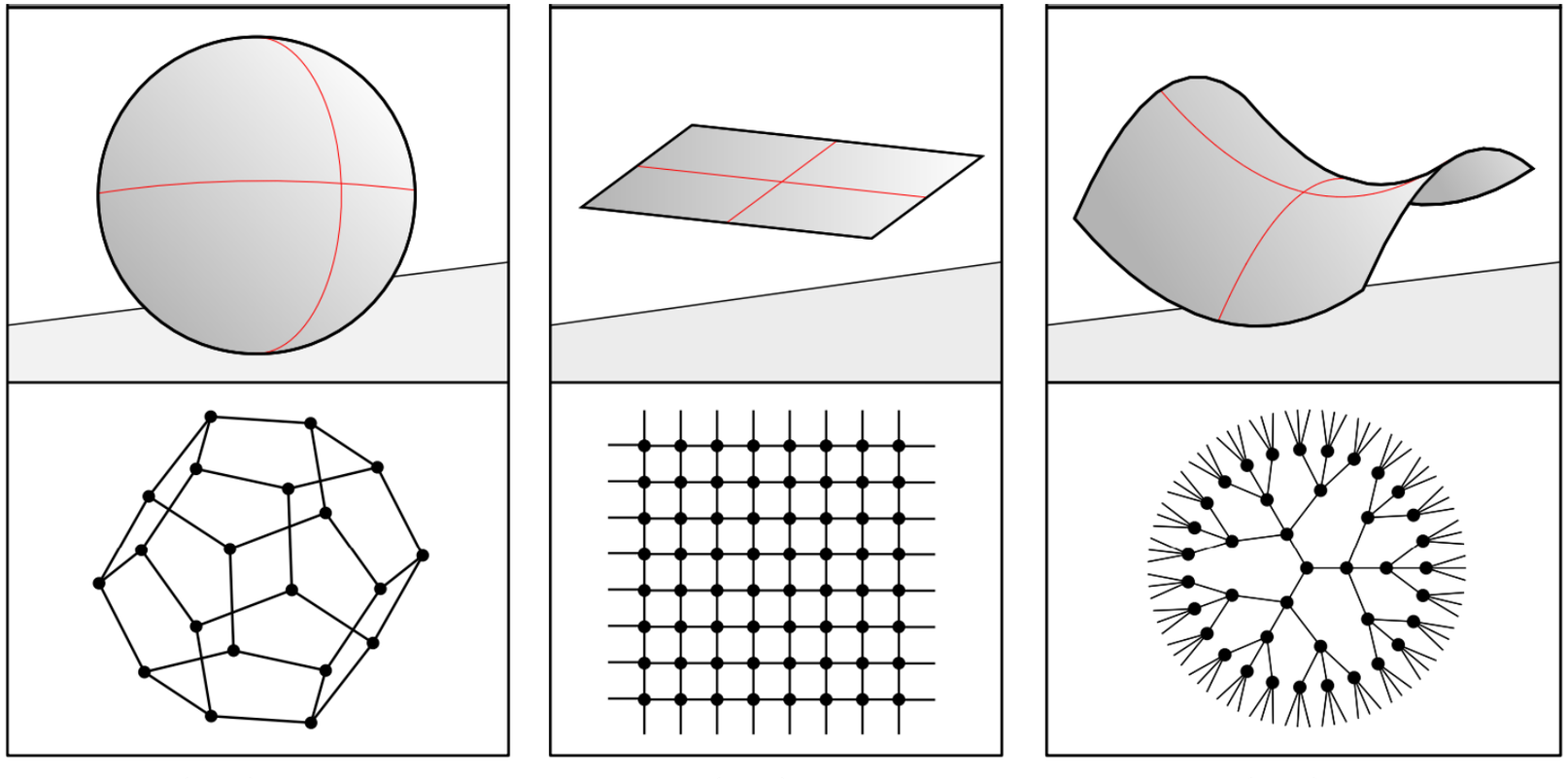
Theorem. Among all manifolds with **Ricci scalar curvature** bounded from below **K < 0**,

$$\Omega^* = \arg \min_{\Omega : R(\Omega) \geq K} \mathbb{E}_{j \in \partial B_\Omega}[\sigma(i_0, j)]$$

has **constant negative curvature K.**

Manifold

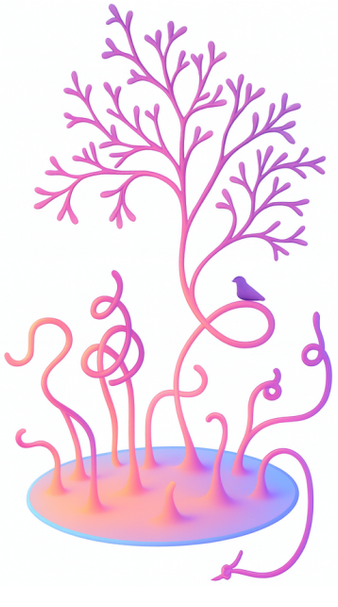
Graph



$R(\Omega) > 0$ Positive curvature

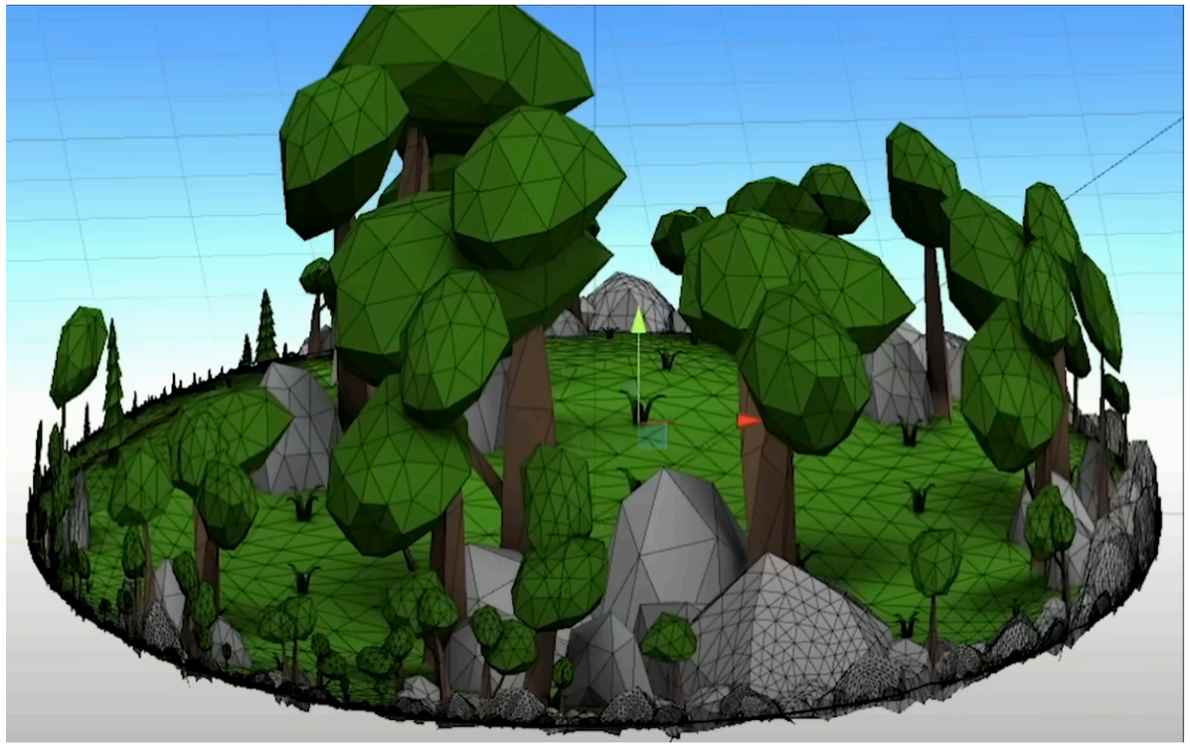
$R(\Omega) = 0$ Zero curvature

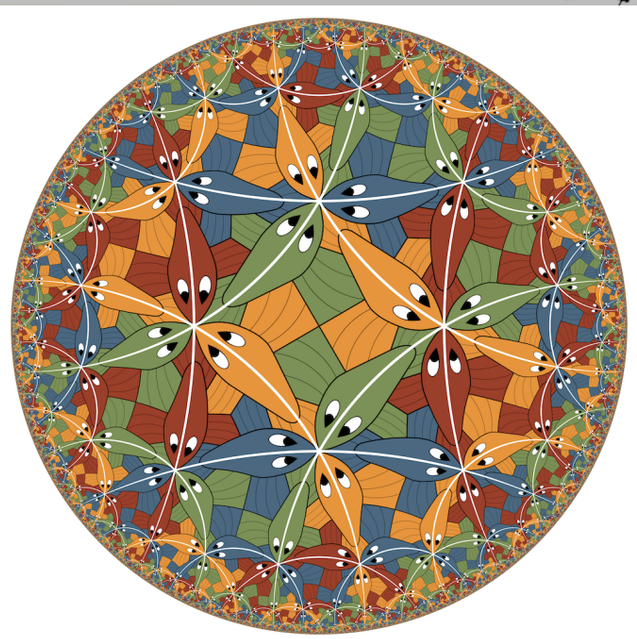
$R(\Omega) < 0$ Negative curvature

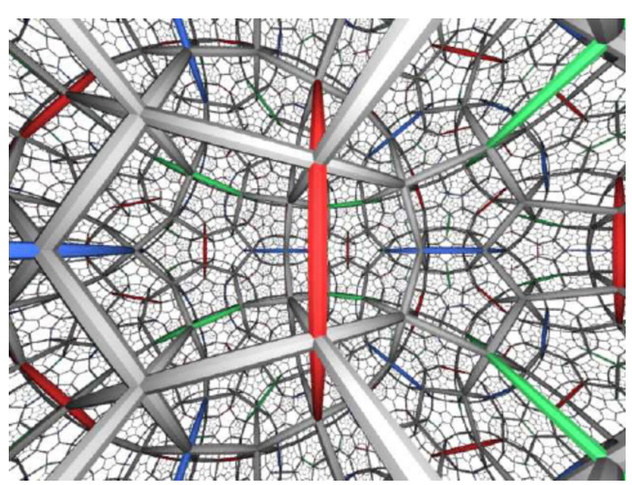


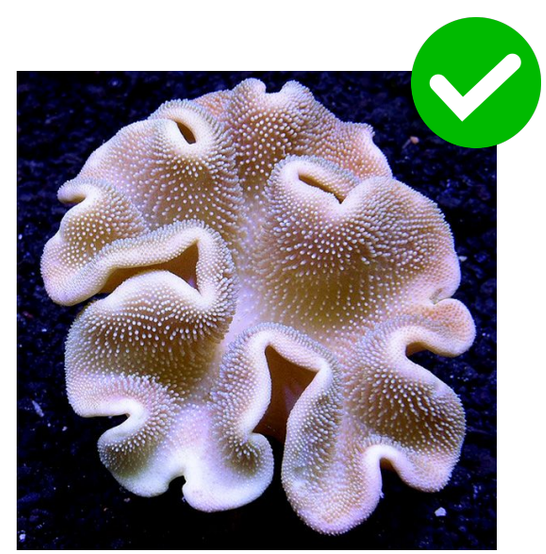
Hyperbolic Space as Optimal Shape

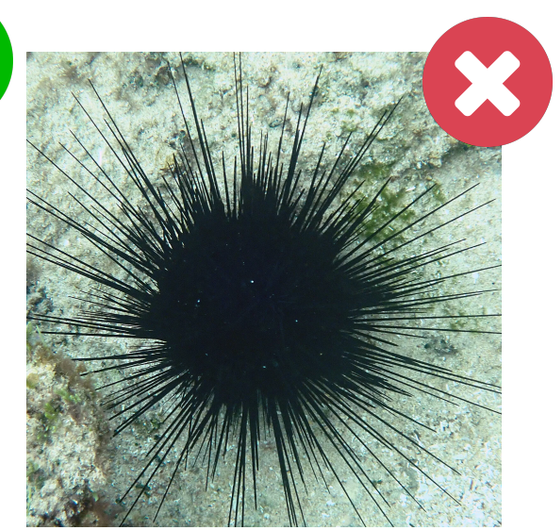
Corollary. The optimal shape to maximize the efficiency of preference inference is **hyperbolic space**.







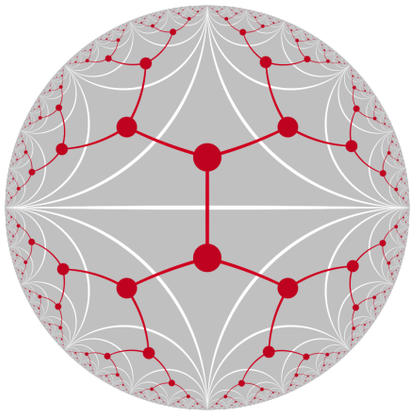


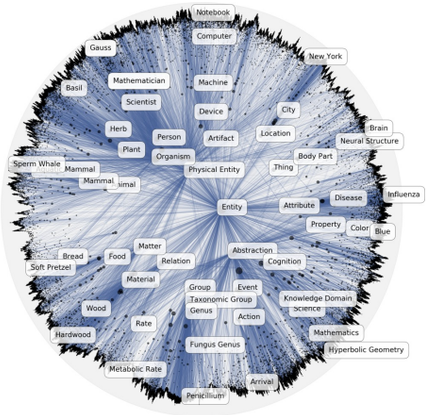


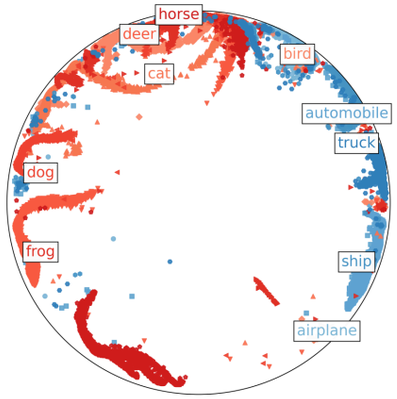
Hyperbolic Space as Optimal Shape

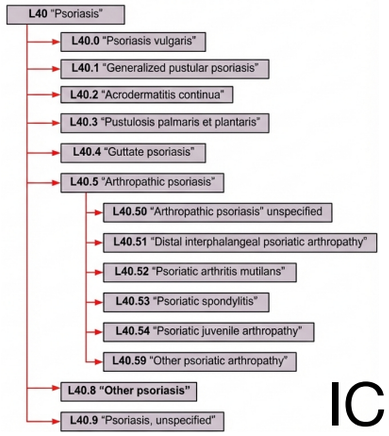
Efficiency optimality matches semantic geometry.

- **Hierarchical structures compatible w. hyperbolic geometry**



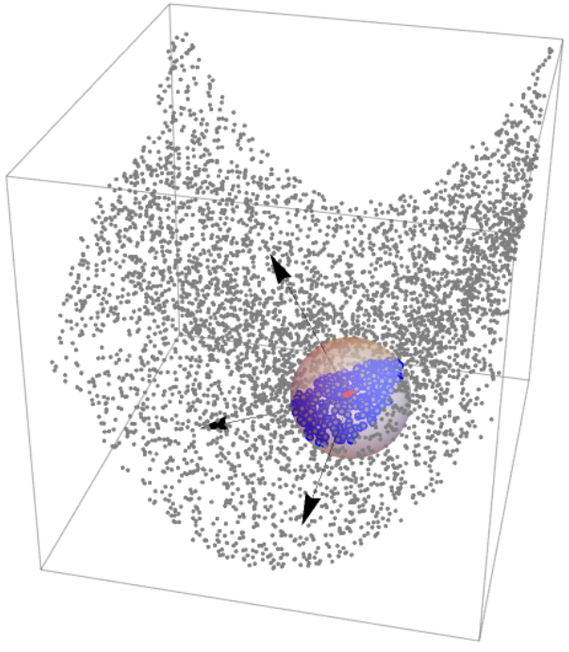


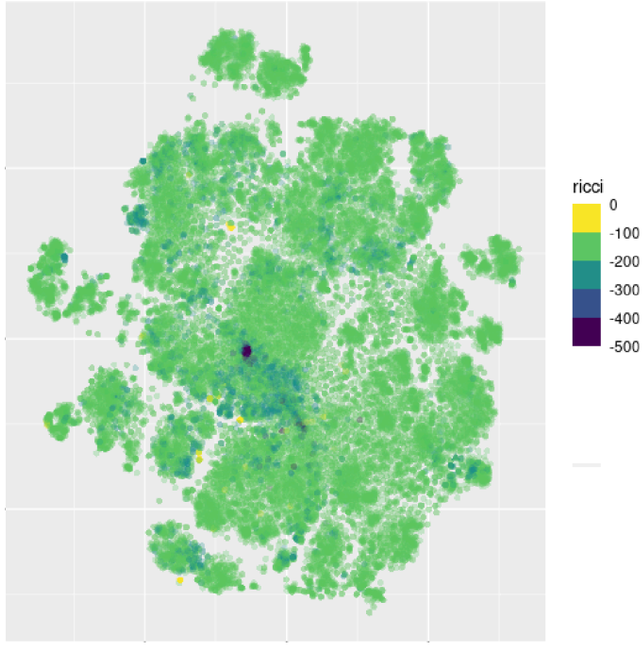




ICD 10

- **Empirical Ricci curvature of Llama2-7B is negative.**
  So just use LLM's own embeddings.





Next Step

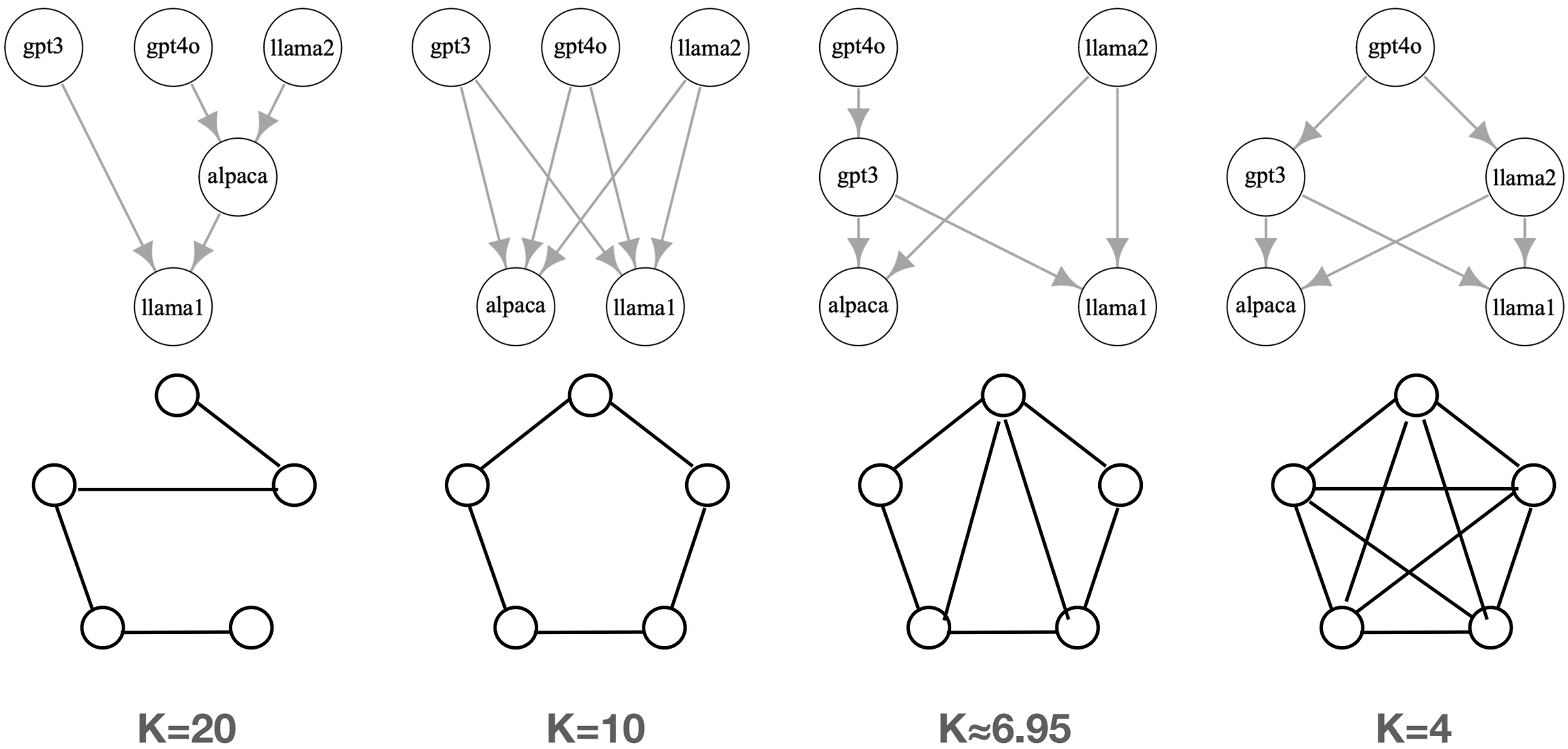
LLM Evaluation: Preference inference for LLM

LLM Fine-tuning: Preference fine-tuning for LLM

• • • ➤ **Application**: LLM for Medical Knowledges

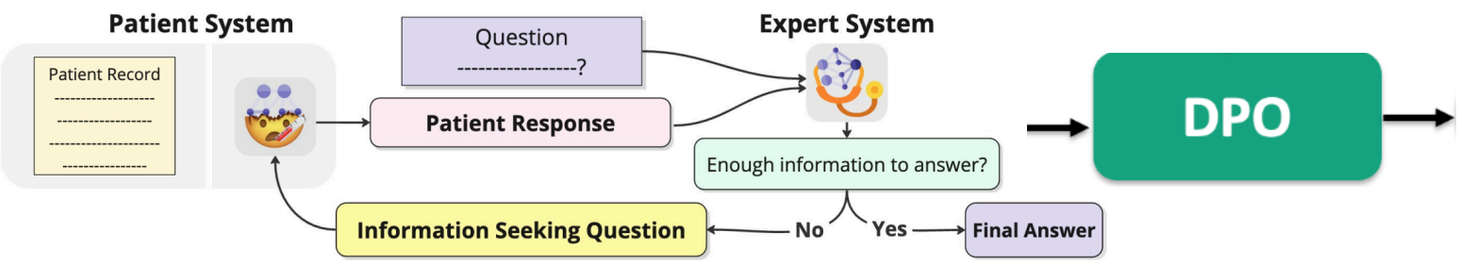
LLM Evaluation vs Comparison Graph

- **Massive Multitask Language Understanding (MMLU) dataset 57 domains; we choose $\Omega =$ medical genetics, LLM-as-judge Sig. level 5%, Kirchhoff index $K = \sum_{i,j} \sigma(i,j)$**



LLM Inference for Medical Questions

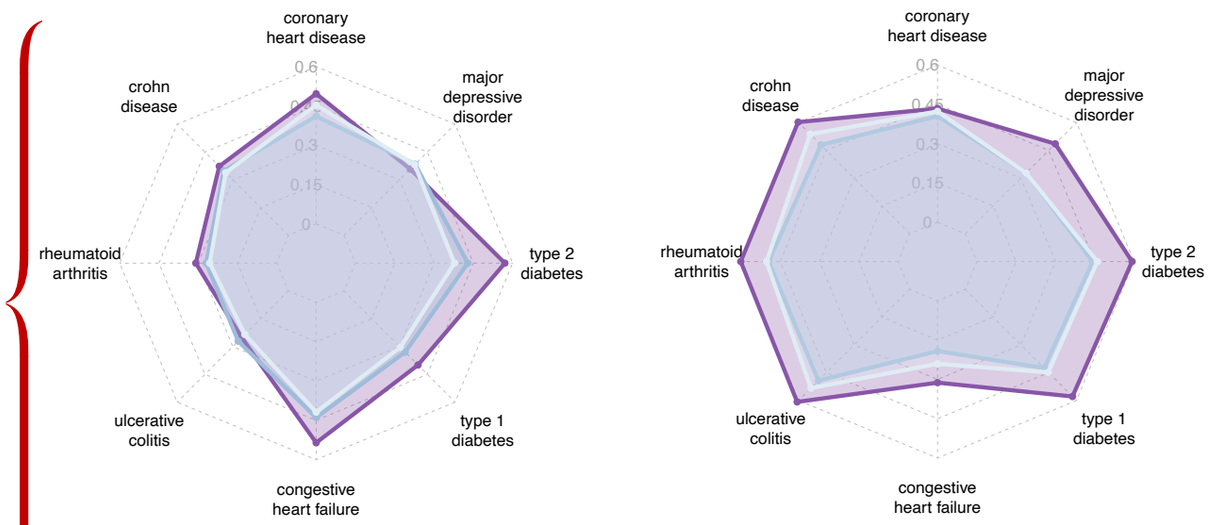
- **MediQ** data: >10k samples of medical QA + preference
- **Task:** Select top answers of $\pi^*(\cdot|\mathbf{x})$



$\mathbf{x}$ = "[Disease] has [relationship] as"

$\mathbf{x}$ = "[Drug] has [side-effects] such as"

Relationships



Comorbidity

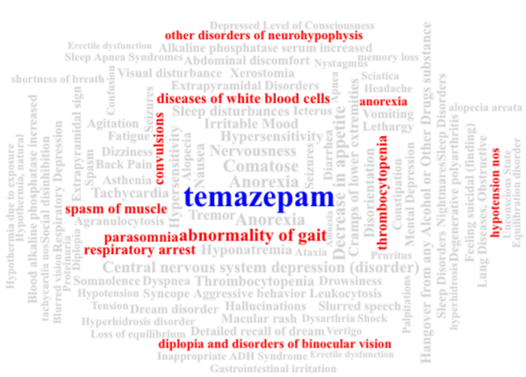
Related drugs

AUC

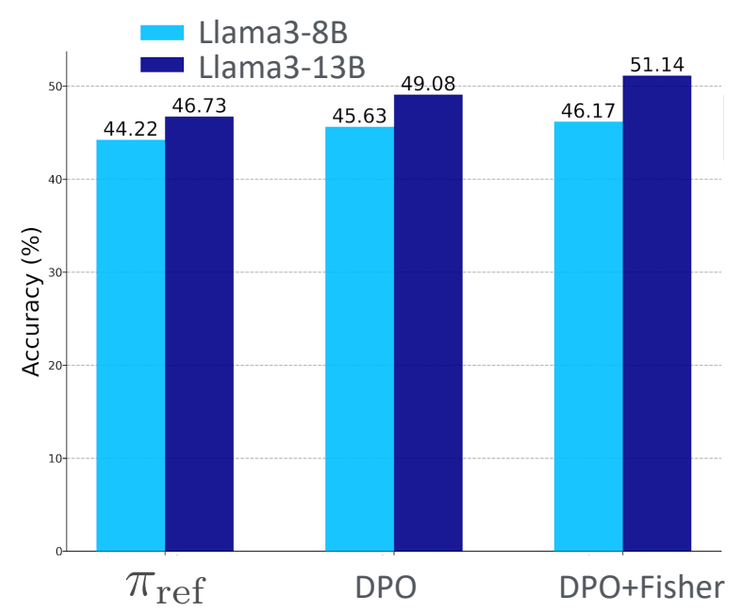
- *DPO+Fisher*
- *DPO*
- π_{ref}*:Llama3-8B*

Side-effects

DPO+Fisher

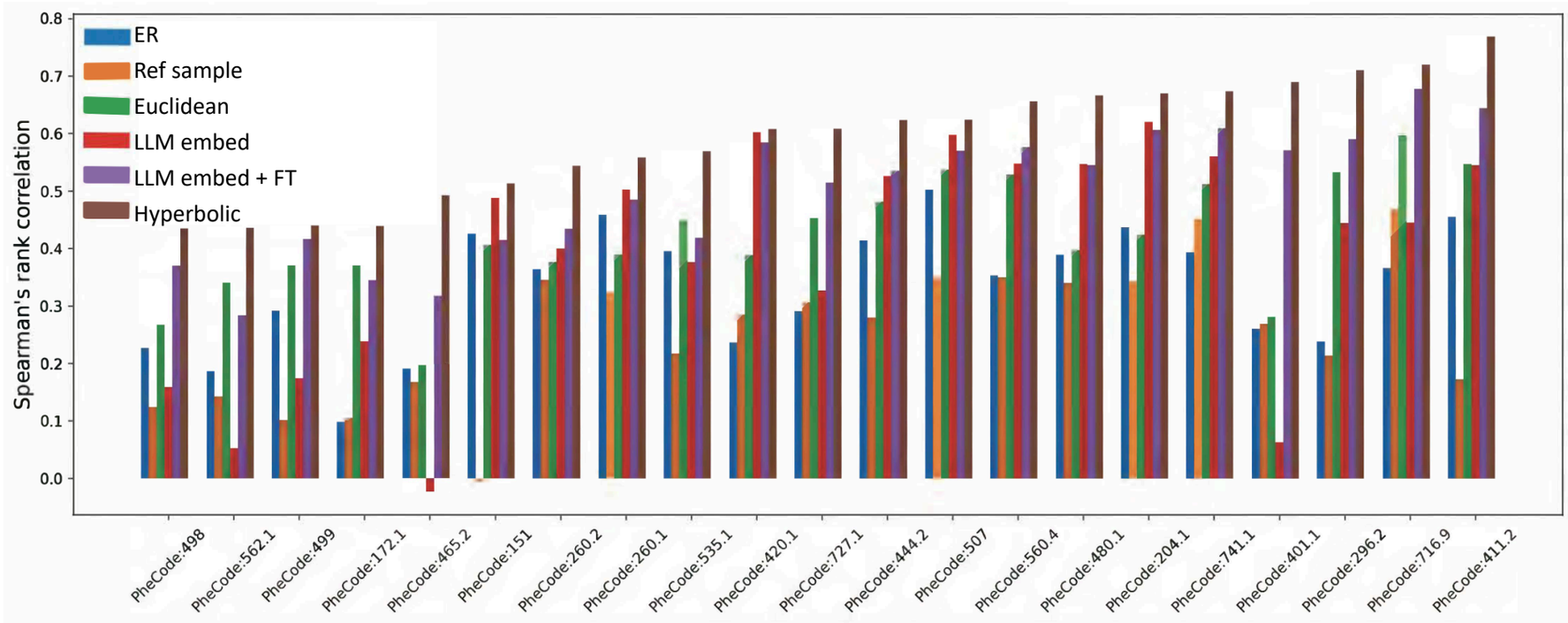


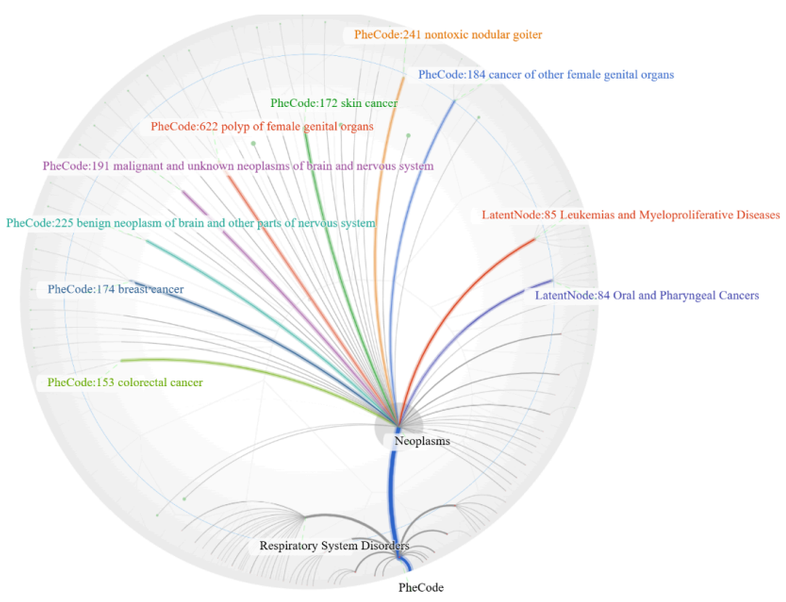
DPO



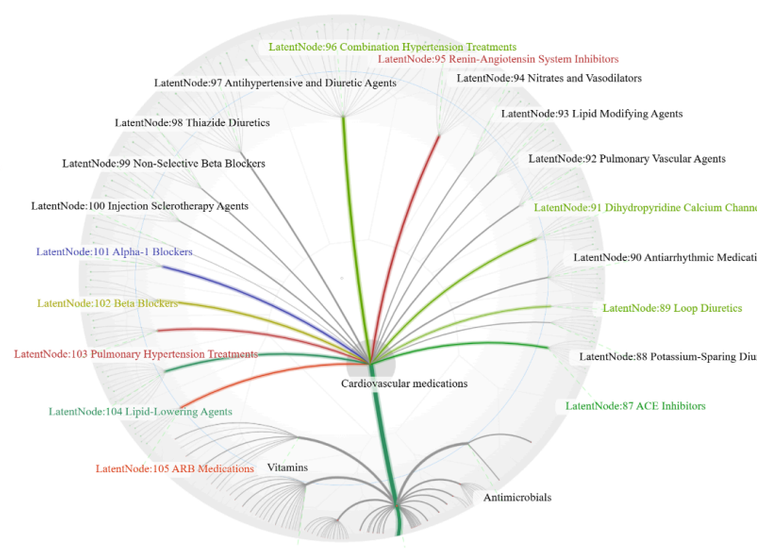
LLM Inference by Shape Design

- **VA+MGB EHR data: 120,032 patients, ~12 yrs window**
- **Task: Predict Phecode relatedness, DPO by diff. sampling**

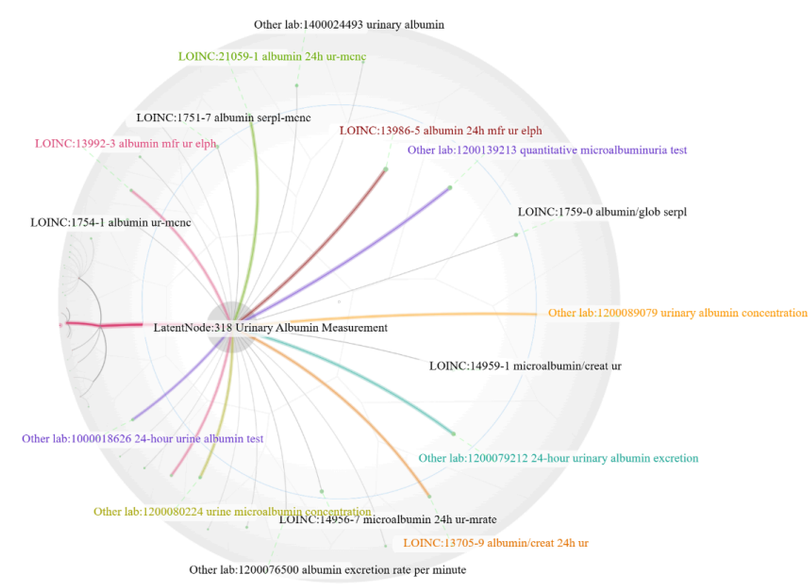




(a) PheCode hierarchy



(b) RxNorm hierarchy



(c) Lab hierarchy

Thank You!