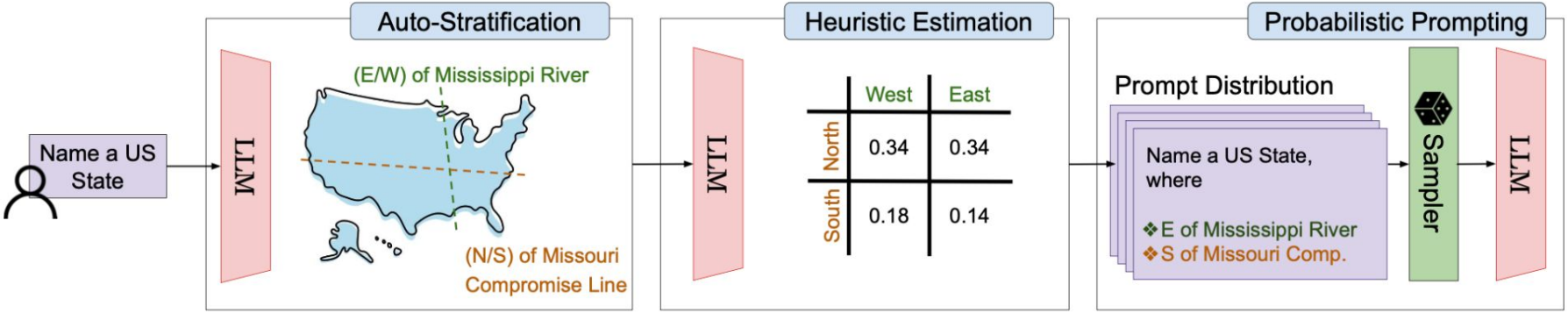
SimpleStrat: Diversifying Language Model Generation with Stratification

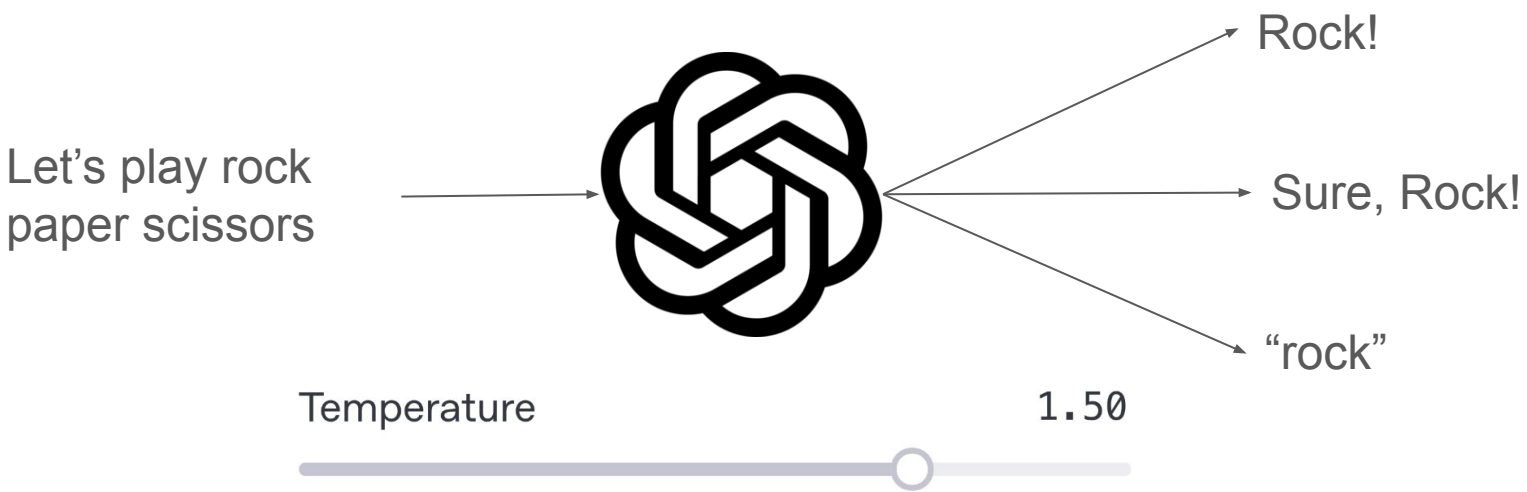


- Propose **SimpleStrat,** a stratified sampling strategy, for improving LLM diversity.
- Introduce **CoverageQA,** a dataset of questions with multiple answers, for measuring diversity.
- We demonstrate as much as **5X improvement to coverage** on CoverageQA and similar pairwise embedding distance at temp 0 as temp 1 on creative writing

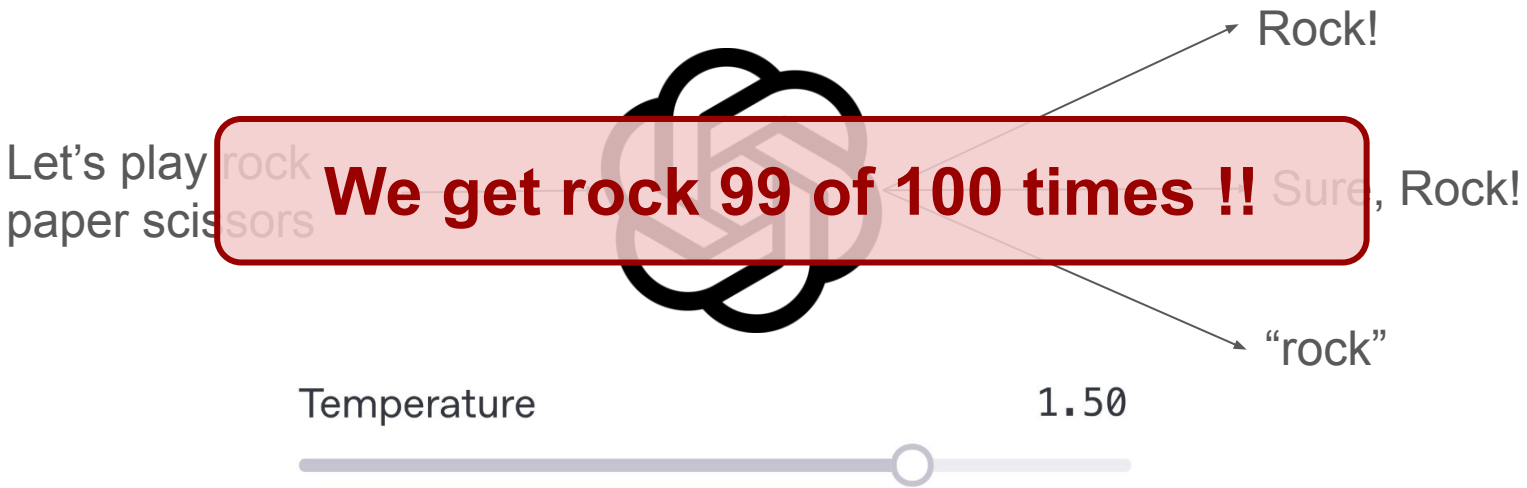
SimpleStrat: Diversifying Language Model Generation with Stratification

Justin W., Yury O., Alex S., Michael L., Sanjit S., and Joseph G.

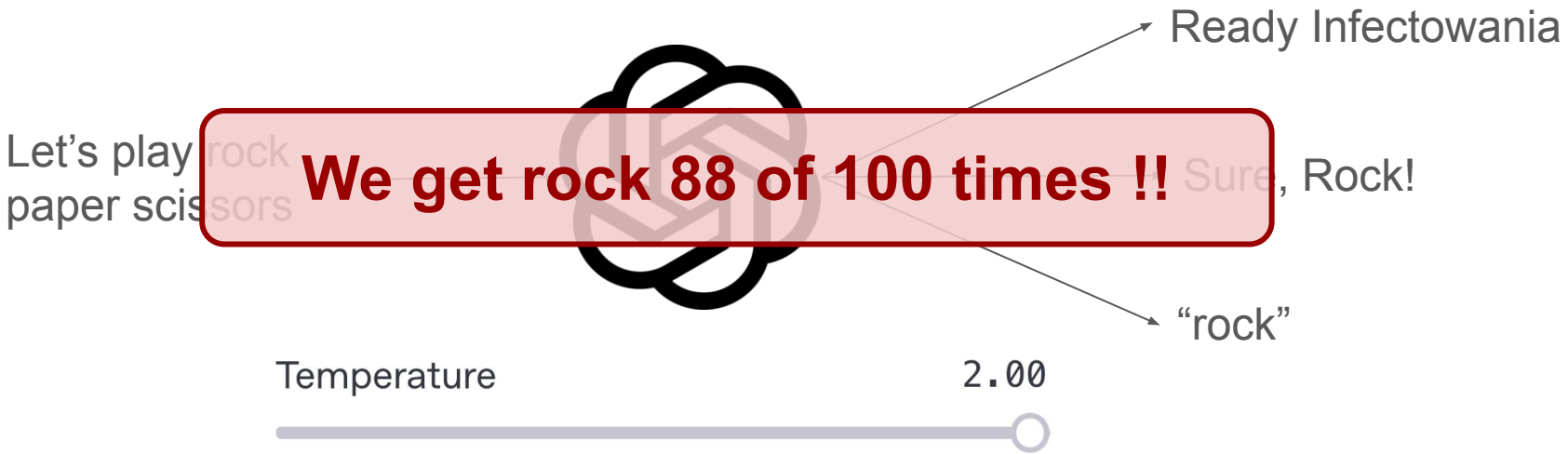
Motivating example:



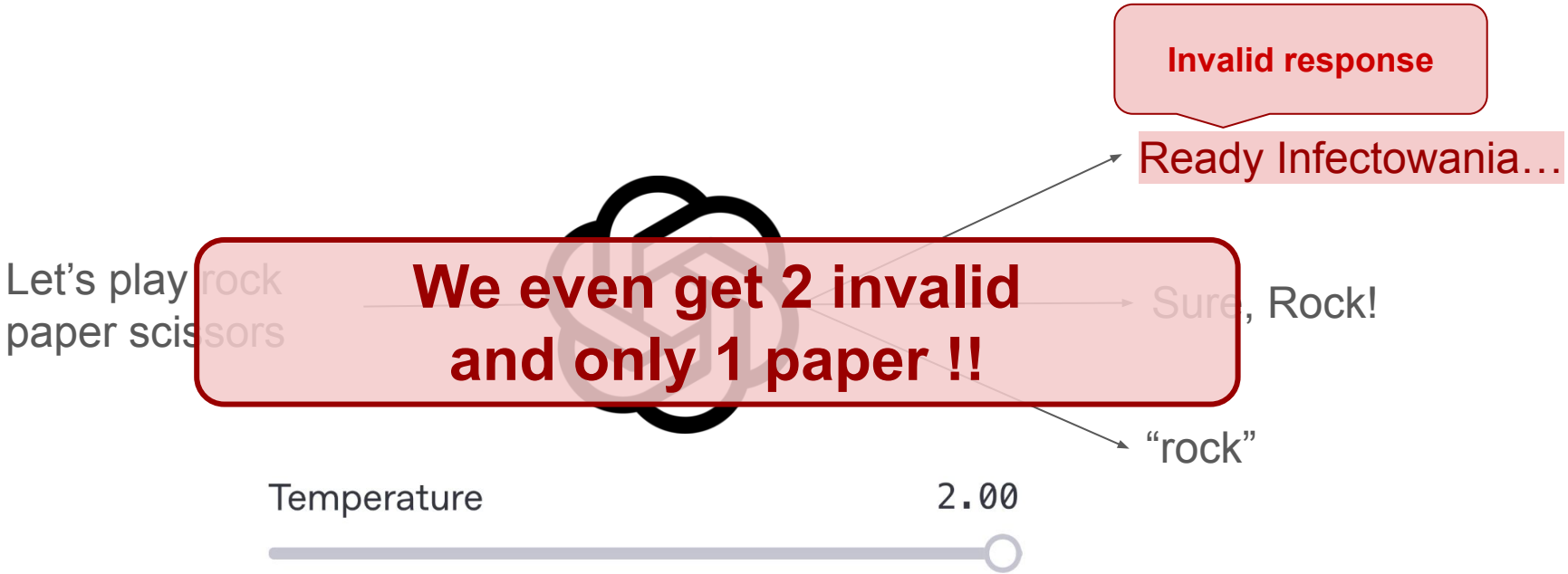
Motivating example:



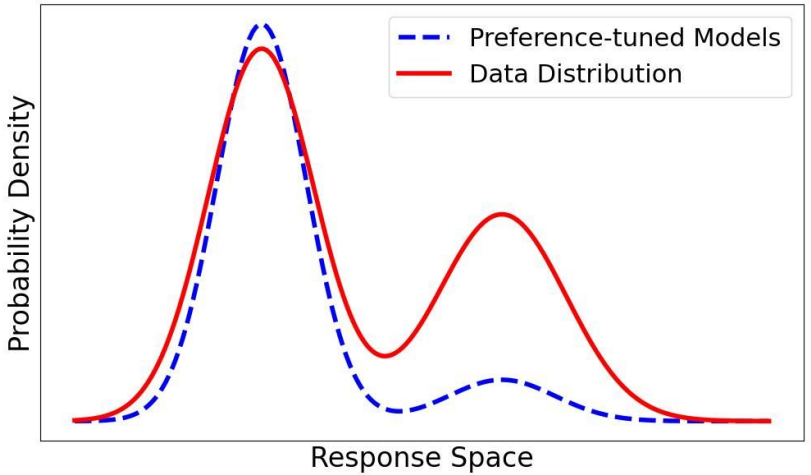
Motivating example:



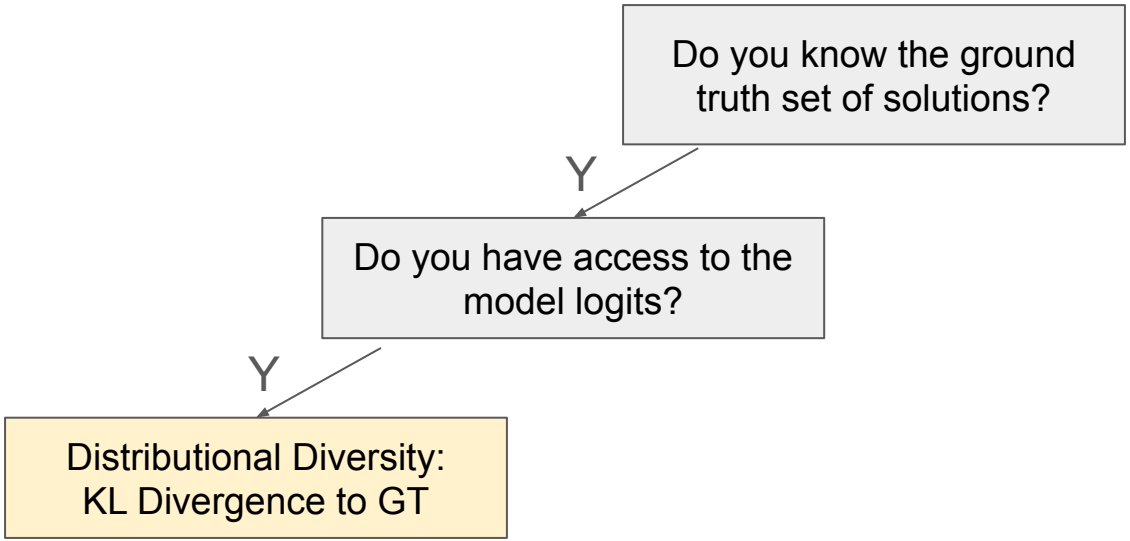
Motivating example:



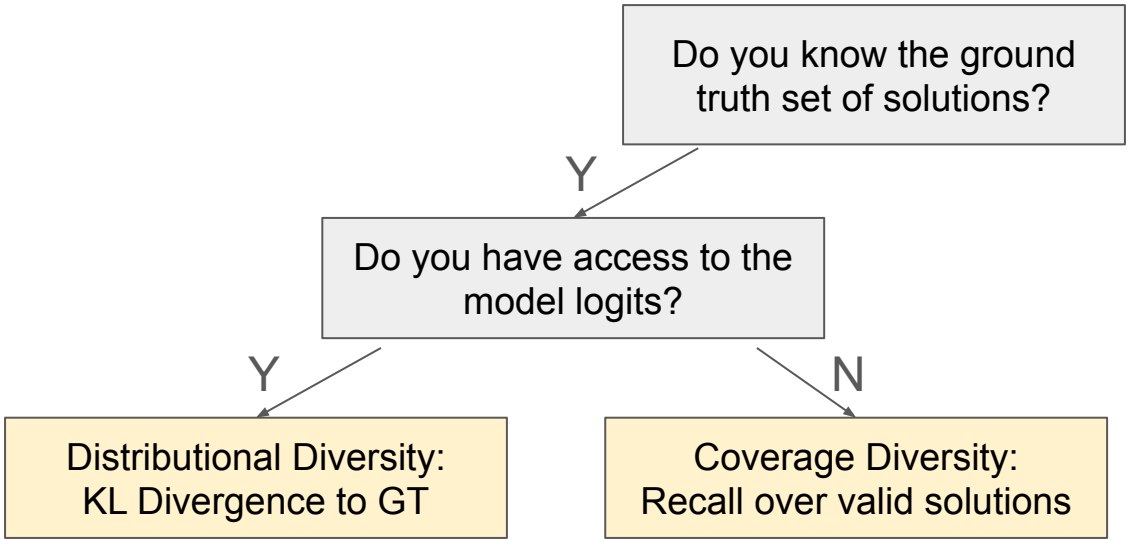
Problem: Language models fail to support modes



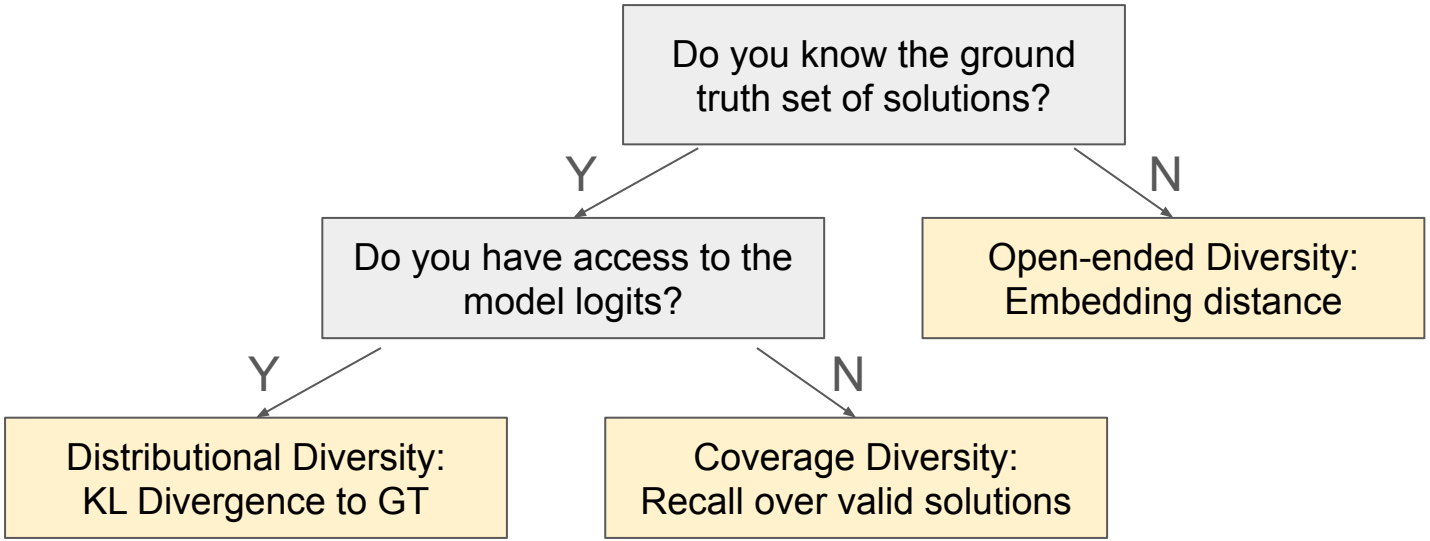
Three Notions of Diversity



Three Notions of Diversity



Three Notions of Diversity



Humans aren’t random

Eyler 2009

We collected data from 119 people who each played 50 games of RPS against a computer playing this optimal RPS strategy. Rather like Kubovy and Psotka’s [4] results, our subjects had a nonuniform preference—for rock. Of the 119 participants, 66 (55.5%) started with rock, 39 (32.8%) with paper, and 14 (11.8%) with scissors.

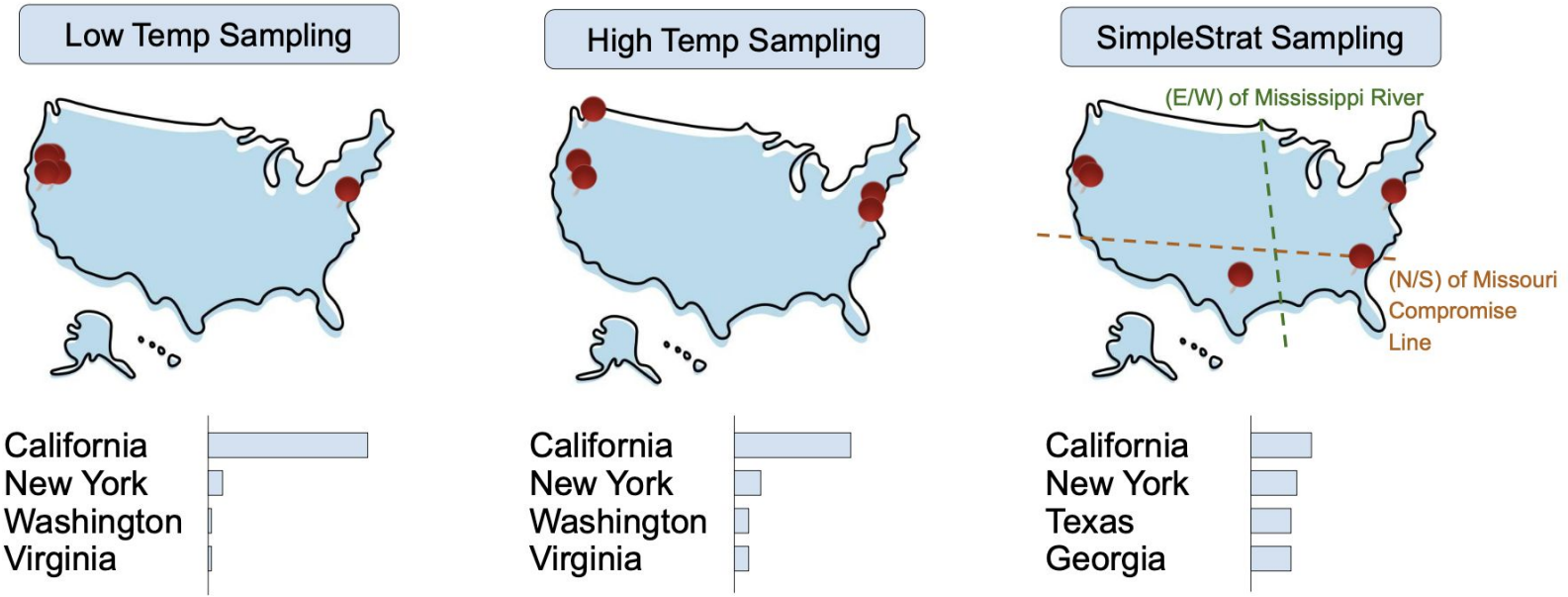
Presidential approval polls

Polls from "select pollsters" meet certain criteria for reliability and are shown with a diamond.

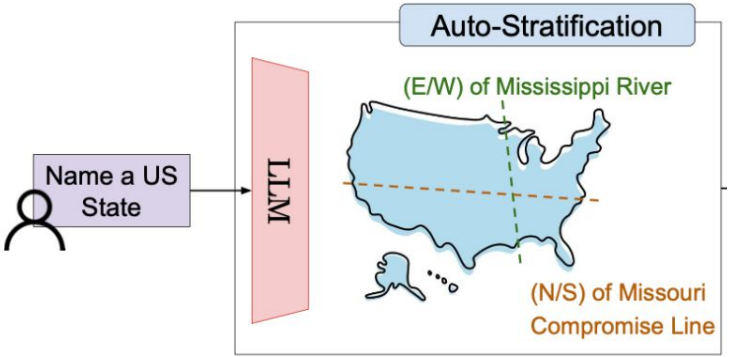
All pollsters | ♦ Select pollsters | Search polls... | Sort by end date

POLLSTER	SPONSOR	NET APPROVAL	APPROVE	DISAPPROVE
Navigator Research Nov. 1-3		Disapprove +15	41%	56%
YouGov Oct. 31 - Nov. 2	Economist	Disapprove +18	39%	57%
Morning Consult Oct. 31 - Nov. 2		Disapprove +7	45%	52%
ActiVote Oct. 1-31		Disapprove +5	46%	51%
TIPP Insights Oct. 28-31	Issues & Insights	Disapprove +11	40%	51%
YouGov Oct. 29-31	CBS News	Disapprove +18	41%	59%
RMG Research Oct. 22-30	Napolitan News Service	Disapprove +1	48%	49%

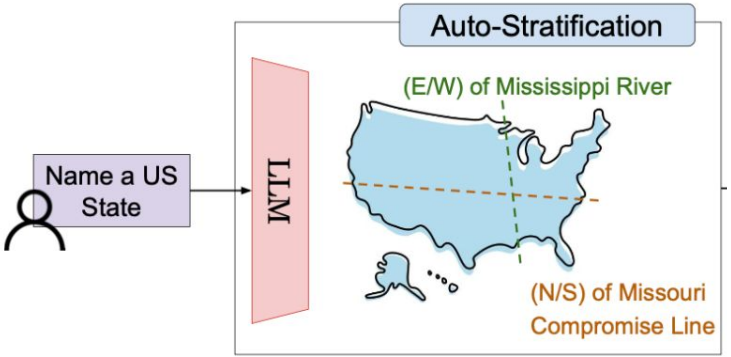
SimpleStrat



Method: Auto-stratification

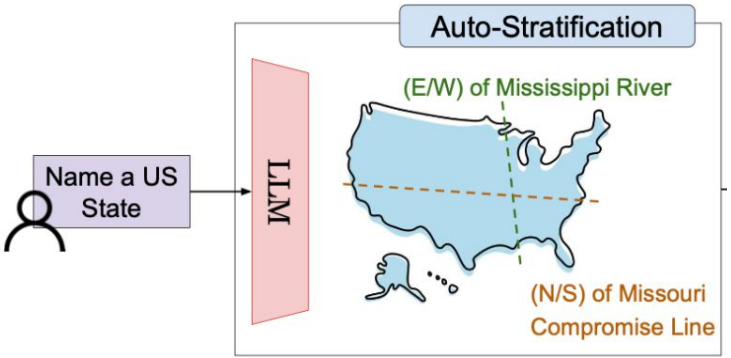


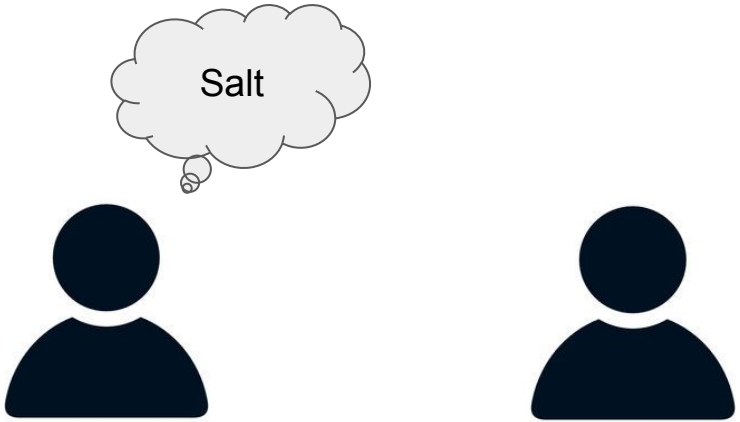
Method: Auto-stratification



20 Questions

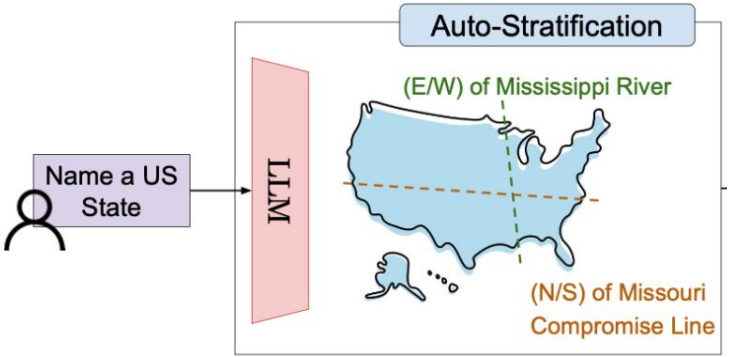
Method: Auto-stratification

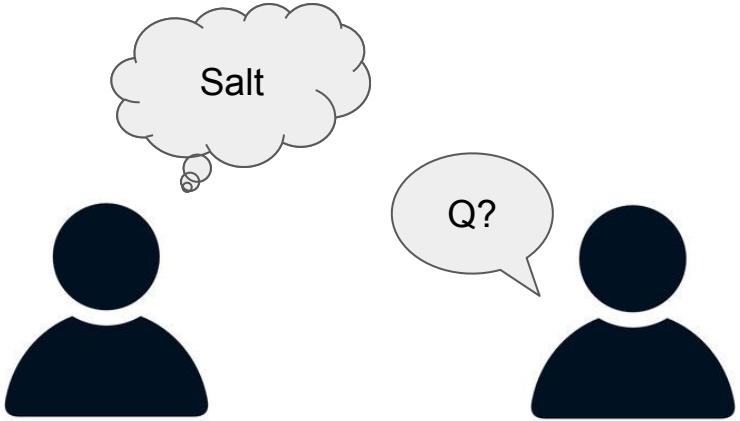




20 Questions

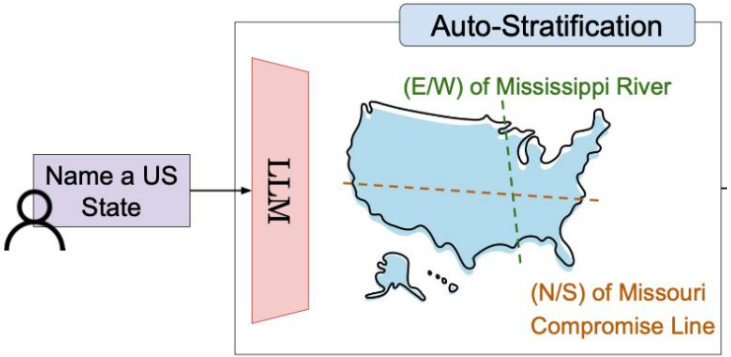
Method: Auto-stratification

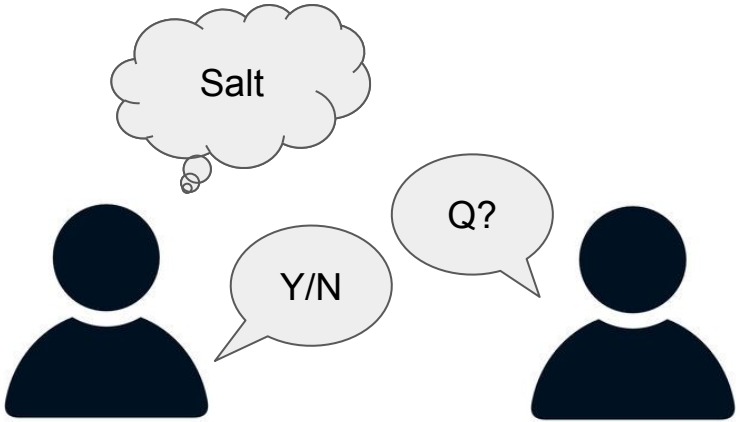




20 Questions

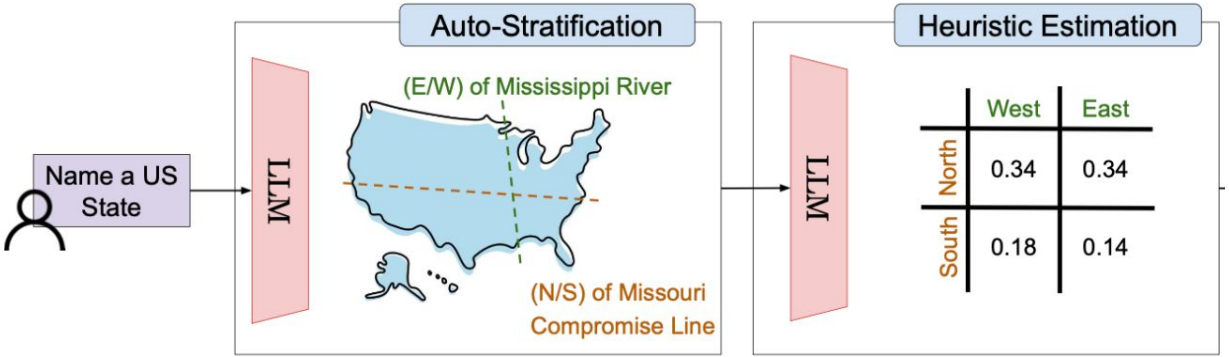
Method: Auto-stratification



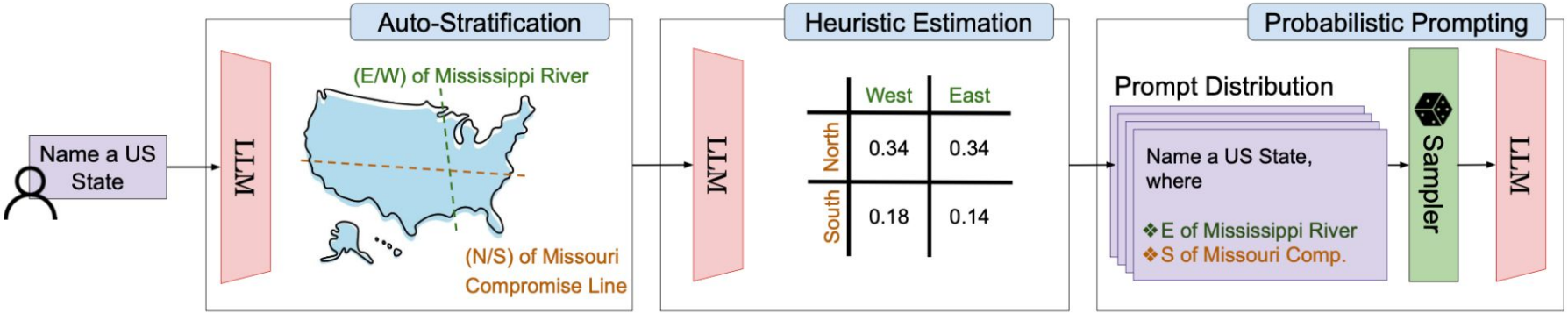


20 Questions

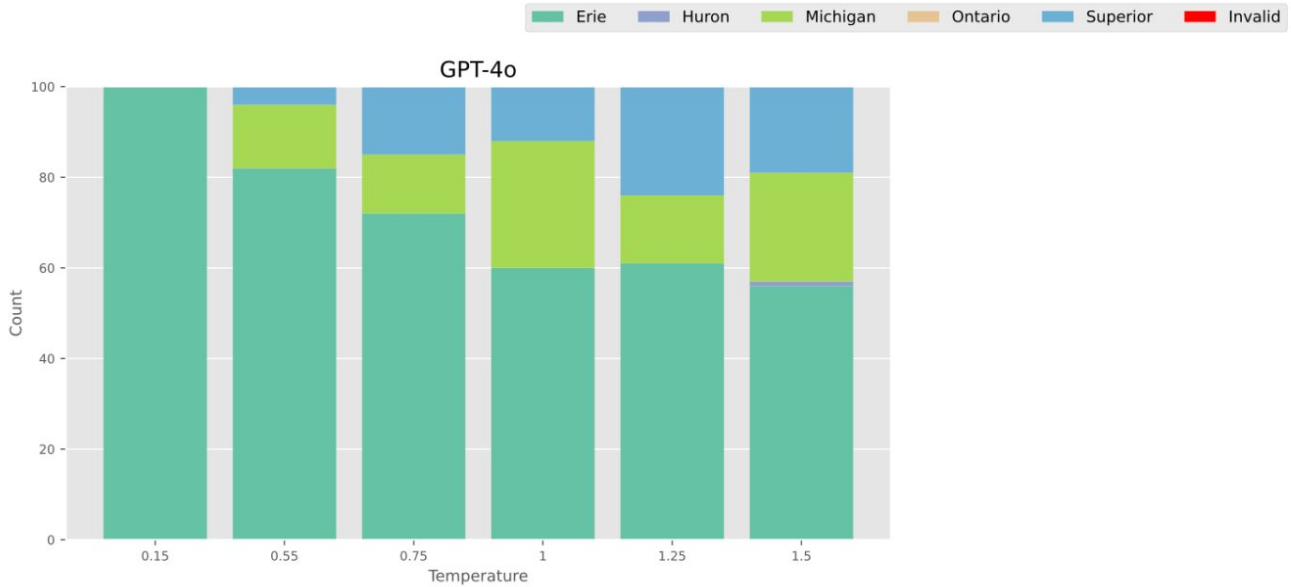
Method: Heuristic Estimation



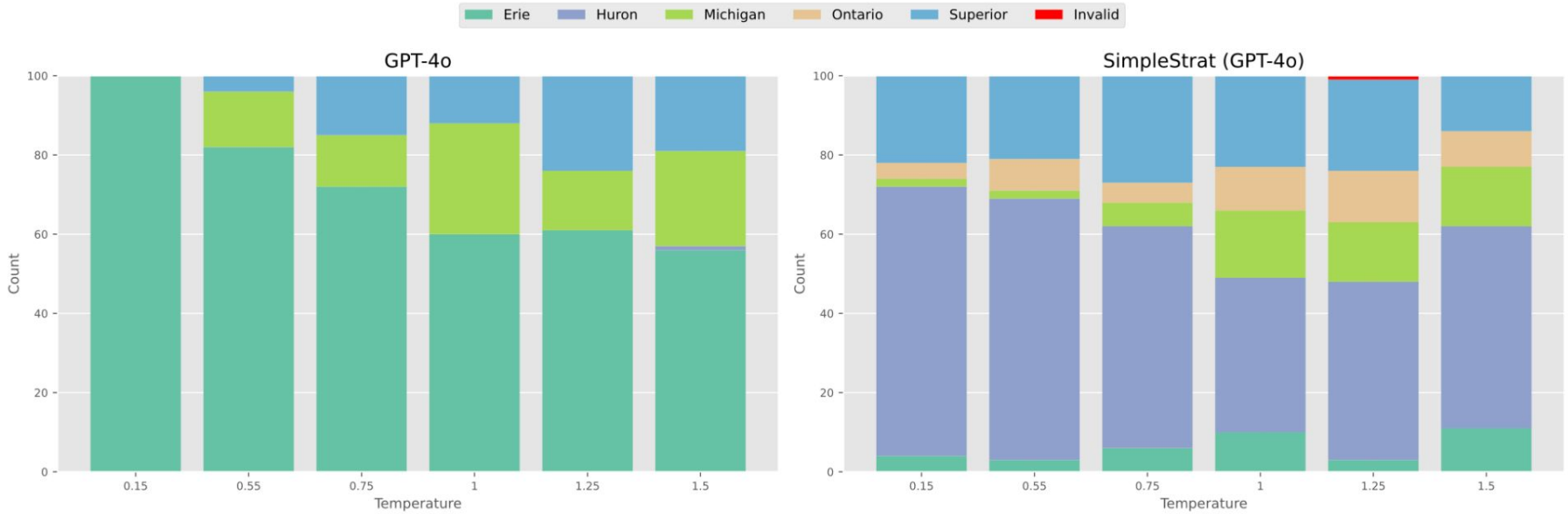
Method: Probabilistic Prompting



Example: “Name a Great Lake”

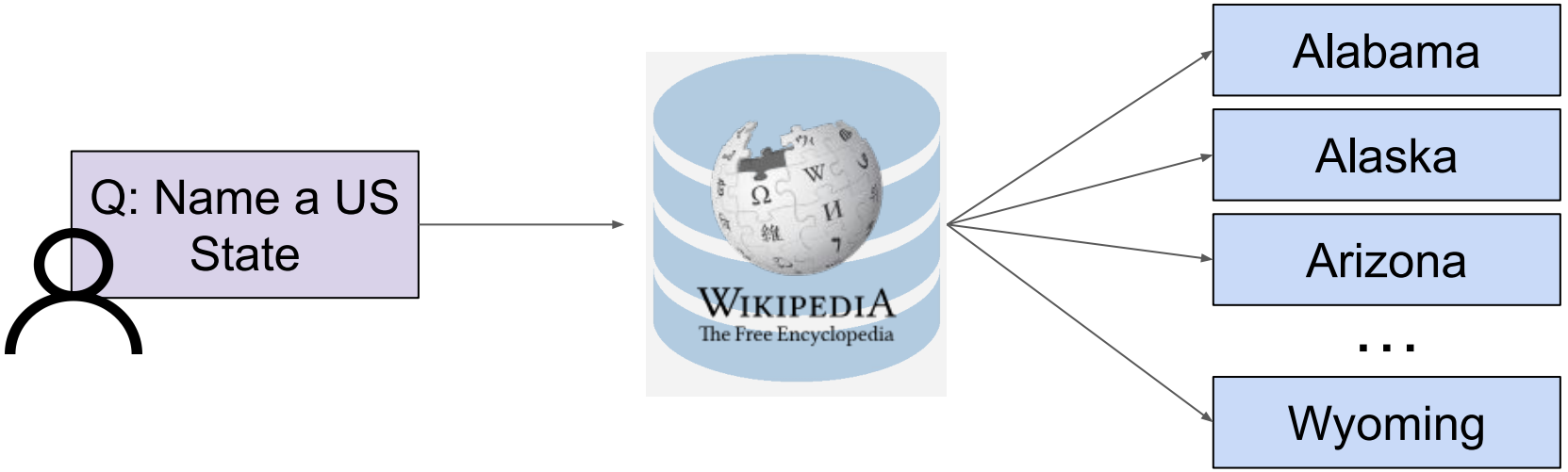


Example: “Name a Great Lake”

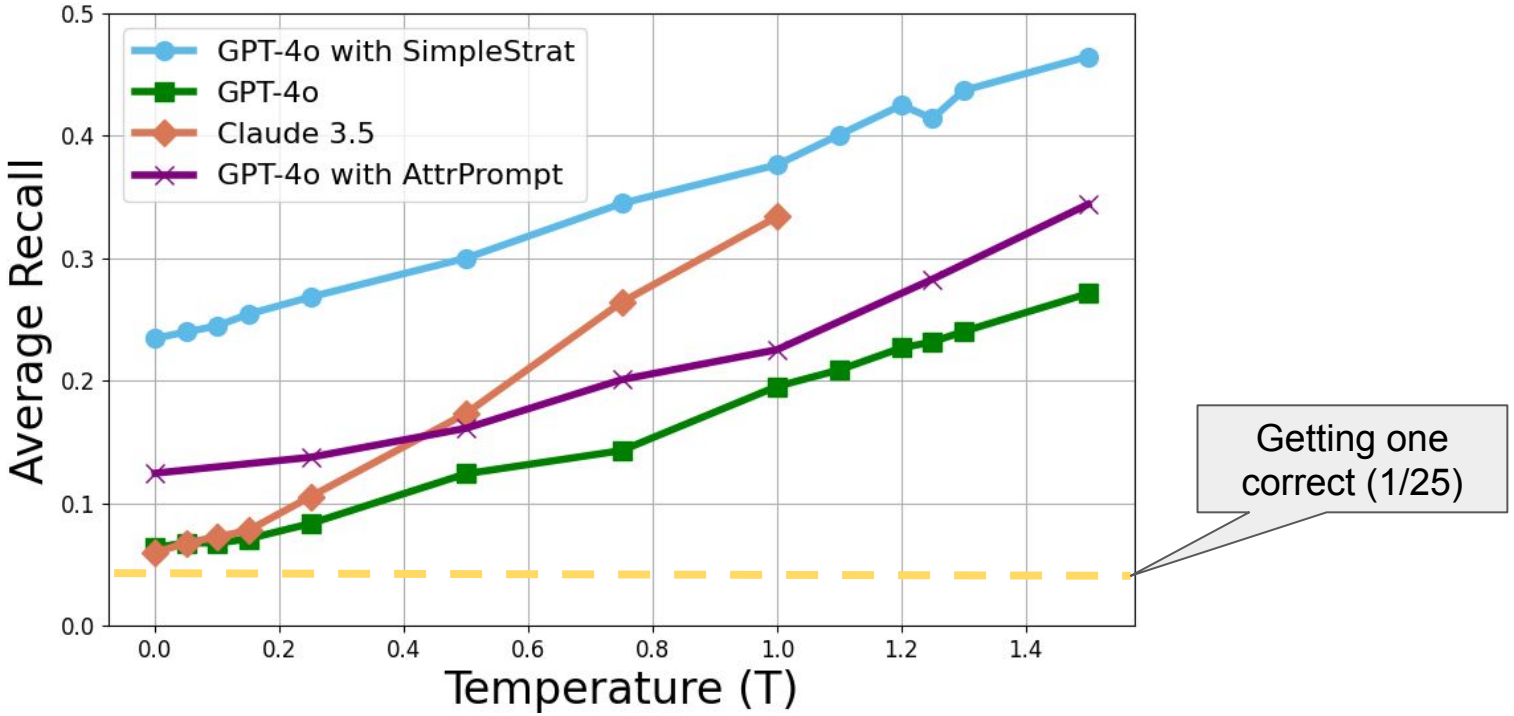


CoverageQA: Multiple correct answers

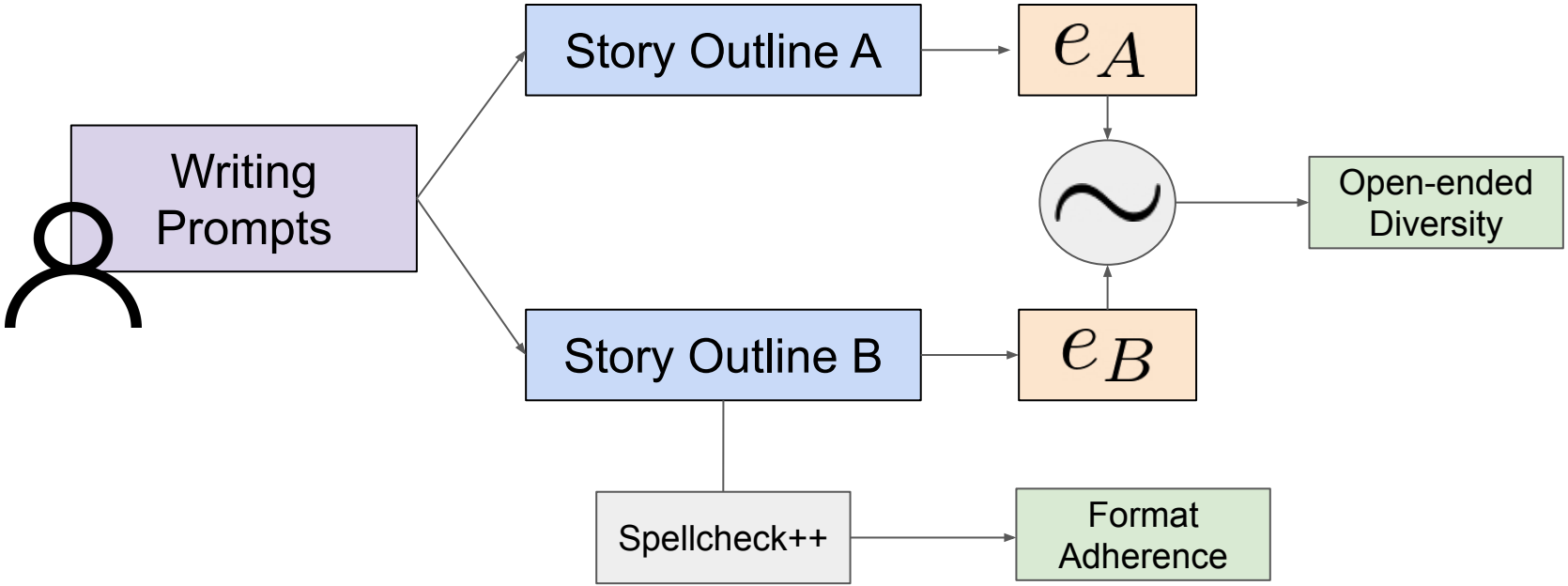
Benchmarks are currently structured to have one correct answer



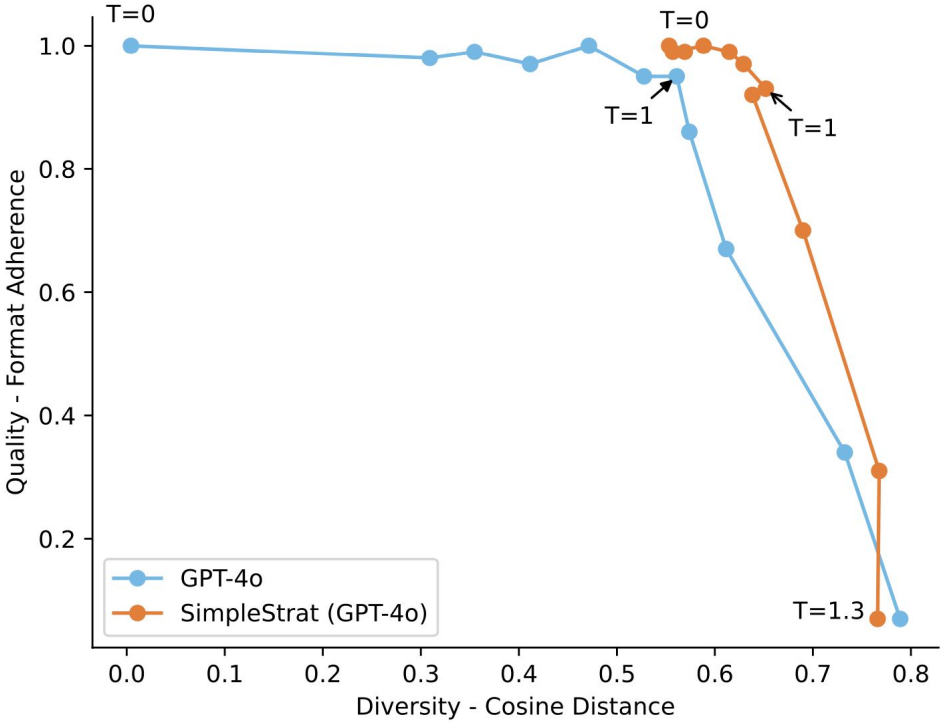
Resampling: Coverage Diversity



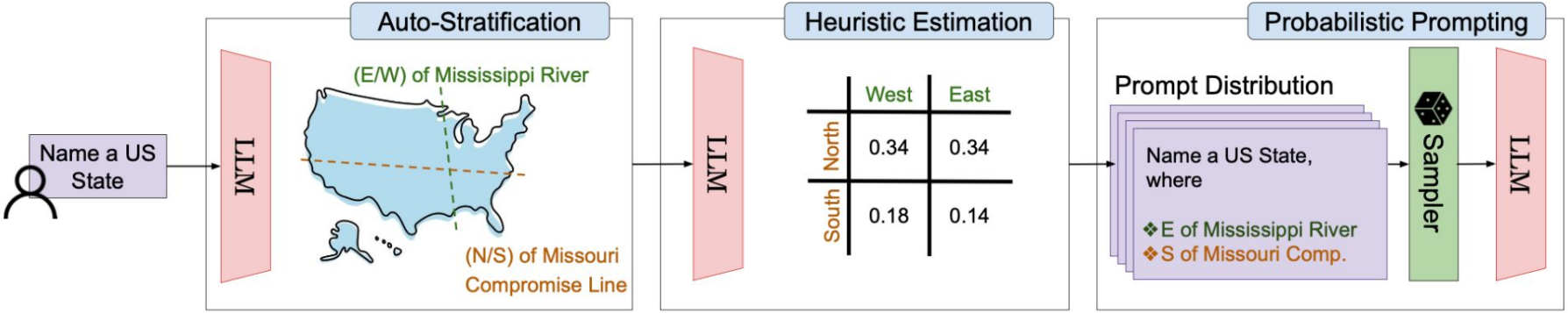
Measuring diversity in creative writing



Story outlines - Open-ended domain



SimpleStrat: Diversifying Language Model Generation with Stratification



- Propose **SimpleStrat,** a stratified sampling strategy, for improving LLM diversity.
- Introduce **CoverageQA,** a dataset of questions with multiple answers, for measuring diversity.
- We demonstrate as much as **5X improvement to coverage** on CoverageQA and similar pairwise embedding distance at temp 0 as temp 1 on creative writing