

## Motivation

- Human Label Variation (HLV) captures genuine annotator disagreement, requiring models to predict *answer distributions*, not only a single label.
- Long Chain-of-Thought (CoT) improves many single-answer reasoning benchmarks, but it may collapse ambiguity into one confident choice.
- Question:** Does CoT help LLMs approximate human label distributions, and if so, what part of the prediction does it actually control?

CoT is a strong *decision maker*, but a weak *distribution calibrator*.



Paper



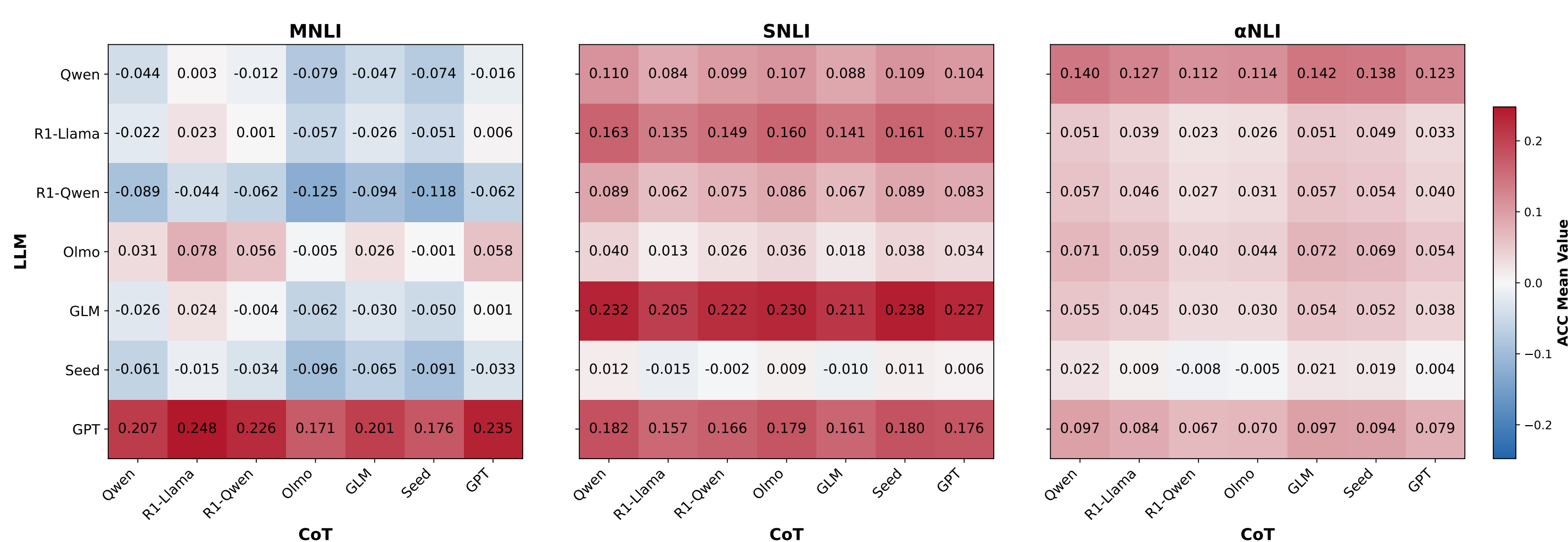
Code

## Experimental Design

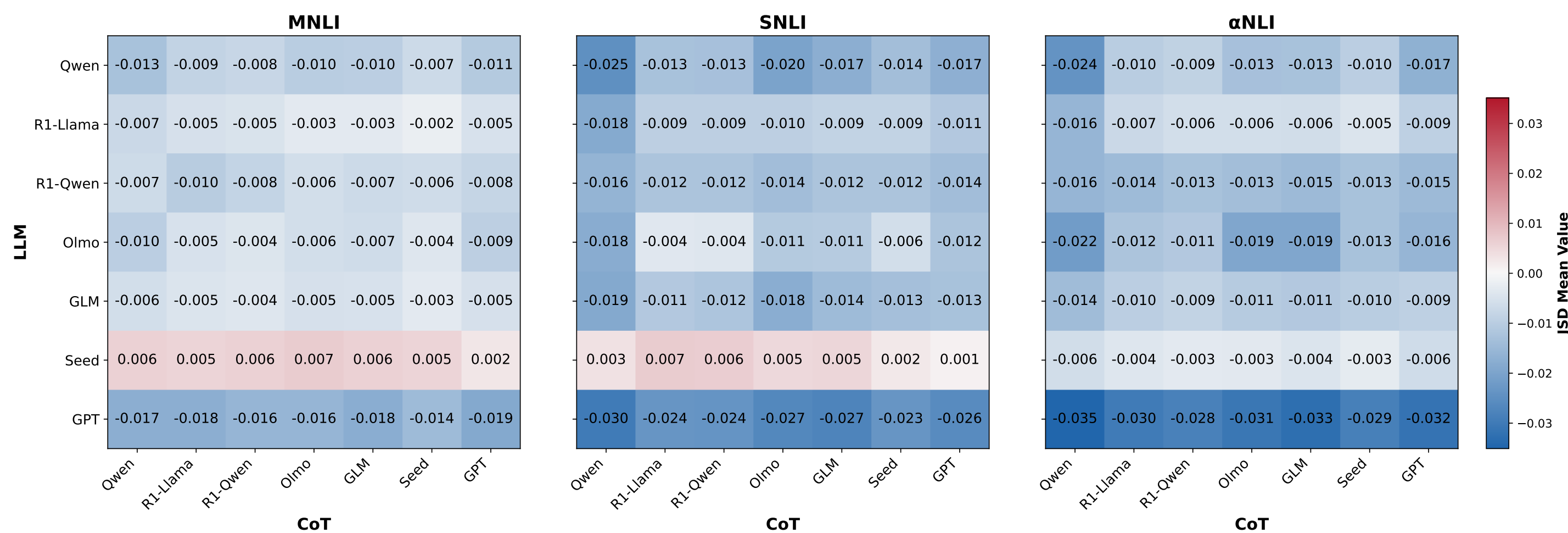
<b>Data</b>	ChaosNLI: MNLI, SNLI, and $\alpha$ NLI, each with human judgment distributions from 100 annotators.
<b>Models</b>	Qwen, R1-Llama, R1-Qwen, Olmo, GLM, Seed, GPT: open-source reasoning-tuned LLMs.
<b>Metrics</b>	Accuracy $\uparrow$ for top-1 correctness; JSD $\downarrow$ for distributional alignment; Spearman's $\rho$ $\uparrow$ for ranking alignment.
<b>Probes</b>	Start-vs-last reasoning, Cross-CoT transfer, and step-wise early stopping over 10 CoT segments.

## Does CoT Improve HLV Performance?

Dataset	Accuracy after CoT	JSD after CoT	Main observation
MNLI	mixed	lower for 6/7 models	distribution improves even when accuracy drops
SNLI	higher for 7/7 models	lower for 6/7 models	CoT helps both top-1 and distributional fit
$\alpha$ NLI	higher for 7/7 models	lower for 7/7 models	strongest overall benefit



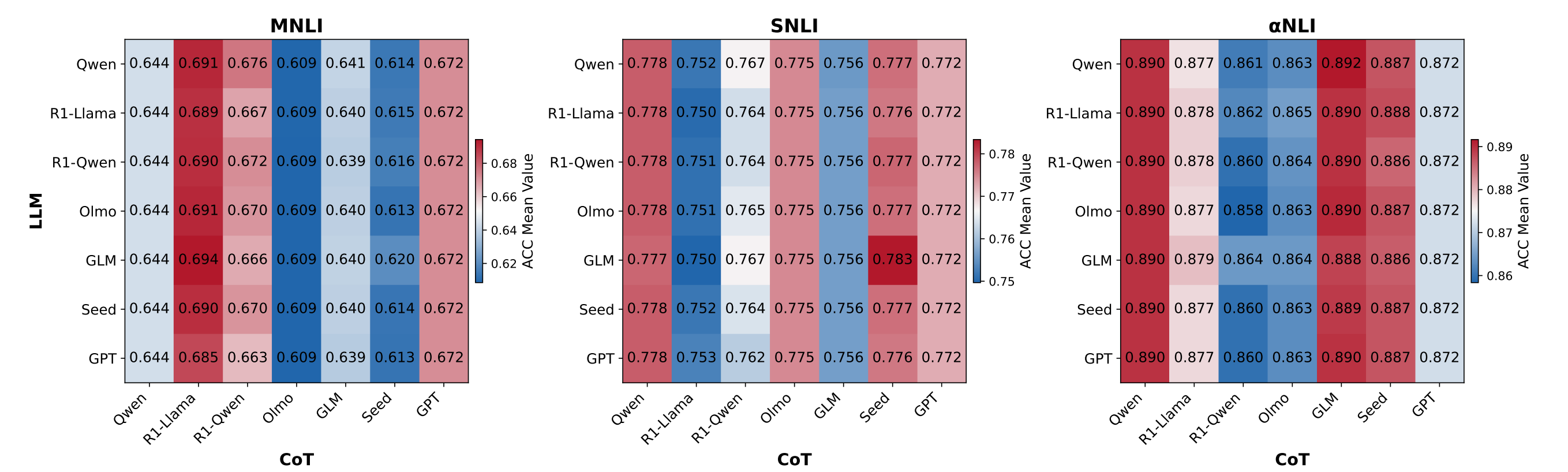
Change in accuracy CoT under Cross-CoT.



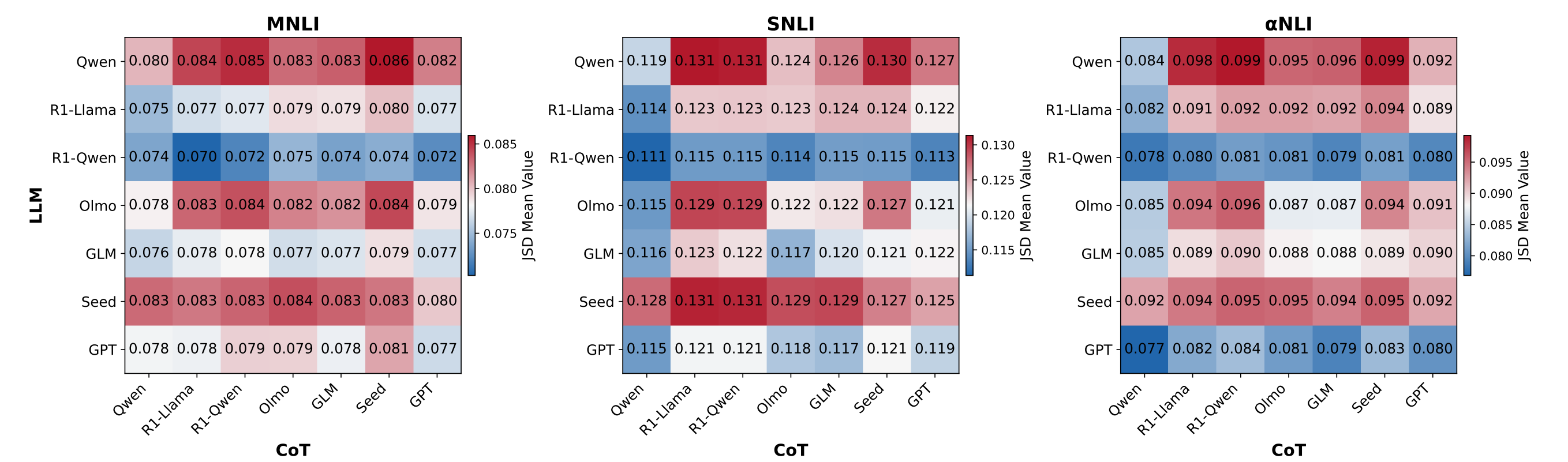
Change in JSD under Cross-CoT.

Injecting CoT almost universally reduces divergence from human judgment distributions, suggesting that CoT text carries portable HLV-relevant information, but accuracy and JSD respond differently.

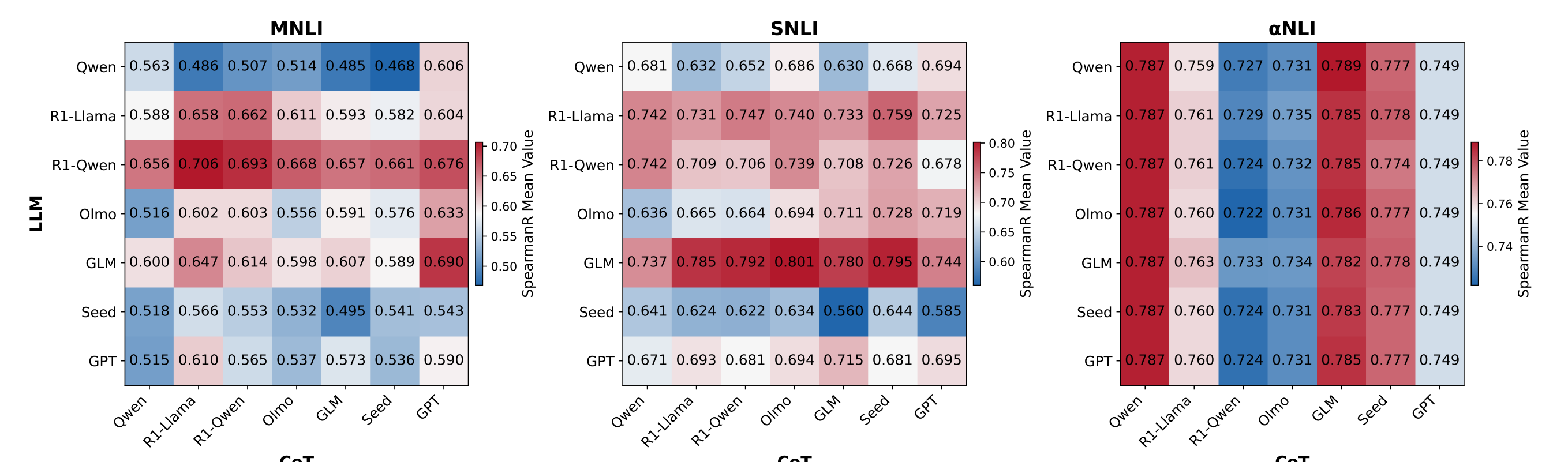
## Cross-CoT Reveals a Split Influence



Final accuracy: column-dominant, mostly controlled by CoT source.



Final JSD: row-dominant, mostly controlled by inference LLM.

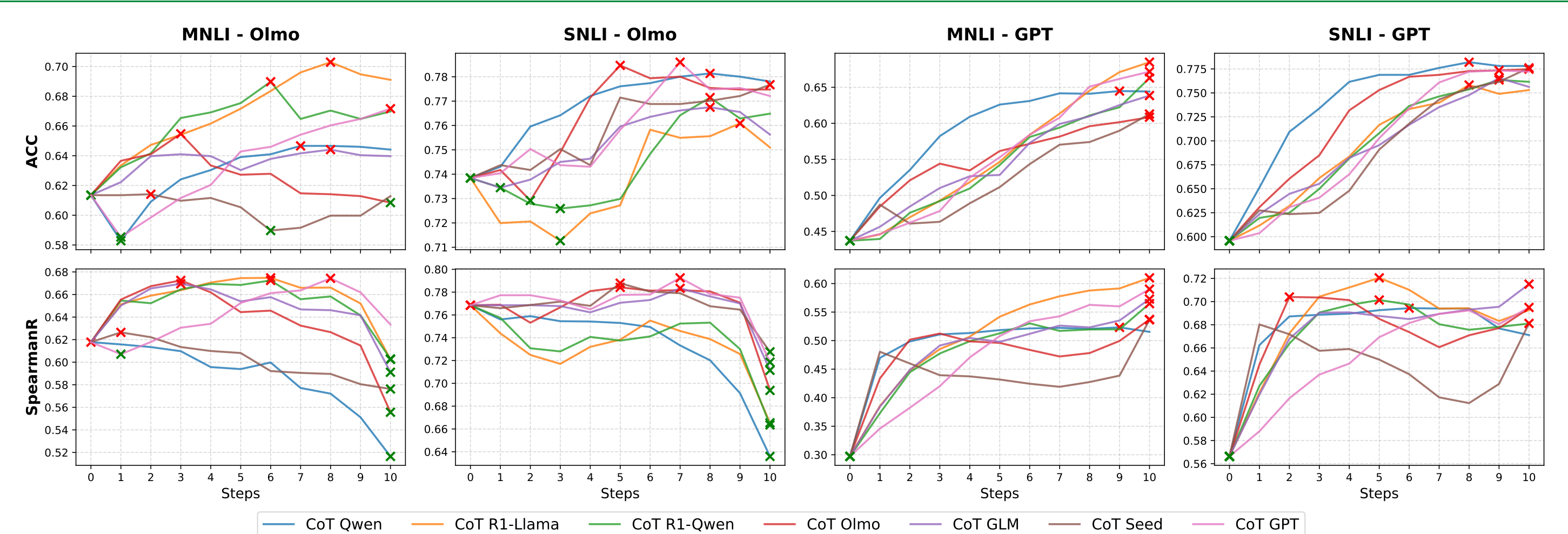


Final ranking alignment: mirrors JSD rather than accuracy.

	MNLI final-step variance LLM prior	CoT	Residual
Accuracy	0.1%	99.5%	0.4%
JSD	83.3%	8.7%	8.1%
Spearman's $\rho$	71.1%	13.1%	15.7%

Models follow CoT for the final argmax choice, but fall back to latent parametric preferences when assigning probability mass to alternative options.

## When Does CoT Take Control?



Representative step-wise curves across reasoning progress.

MNLI step	Accuracy		JSD		Spearman's $\rho$	
	LLM	CoT	LLM	CoT	LLM	CoT
0	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%
5	85.2%	4.3%	89.3%	4.2%	88.0%	4.9%
10	0.1%	99.5%	83.3%	8.7%	71.1%	13.1%

CoT influence on accuracy spikes at the final reasoning step, while its influence on distributional similarity remains low throughout the chain.

## Conclusion

- Long CoT improves overall distributional alignment, suggesting that reasoning text does contain useful information for HLV.
- But CoT mainly controls the top answer. The non-argmax probability landscape remains anchored to model priors.
- Next step: distribution-aware reasoning. Reasoning systems should explicitly maintain and calibrate uncertainty instead of only converging to a decisive label.