

Landscape of Thoughts: Visualizing the Reasoning Process of Large Language Models

Zhanke Zhou

joint work with Zhaocheng Zhu, Xuan Li, Mikhail Galkin, Xiao Feng, Sanmi Koyejo, Jian Tang, Bo Han

TLDR: We introduce the first visualization tool to inspect the LLMs' reasoning trajectories

paper: <https://arxiv.org/abs/2503.22165>

code: <https://github.com/tmlr-group/landscape-of-thoughts>

homepage: <https://landscape-of-thoughts.github.io/>

Motivation: Why visualize?

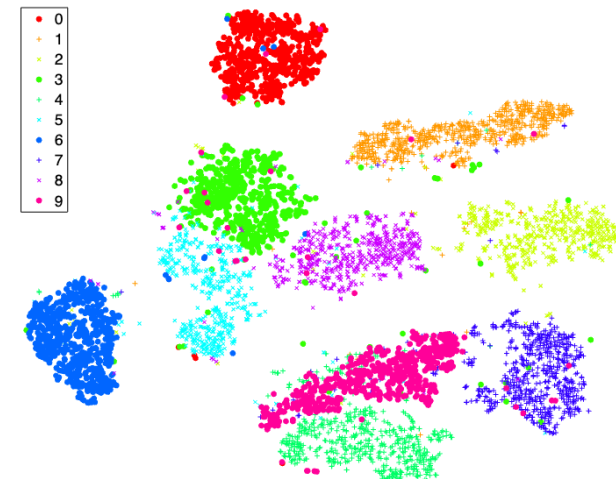
The problem states that a car is being driven towards the base of a

Here are the step-by-step thoughts for each question: Question 1:

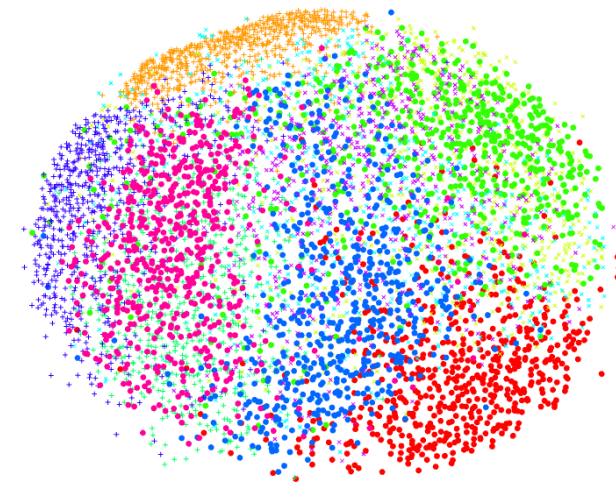
To solve the problem systematically, let's proceed step by step. The scenario involves a car being driven towards a tower, with the initial angle of elevation from the car to the top of the tower being 45° . As the car continues to move and the angle of elevation changes to 60° , the distance to the tower increases. Applying trigonometry, we know that the tangent of the angle of elevation is equal to the ratio of the tower's height (h) to its base distance from the car. Initially, with an angle of 45° , we have $\tan(45^\circ) = h / bT$, which simplifies to $1 = h / bT$. Using the same approach for a new angle of 50° , we get $\tan(50^\circ) = h / b$, or equivalently, $b = h / \tan(50^\circ)$. The car stops when the angle of elevation changes due to the relative motion with respect to the tower's top, but since we're told the full process takes 10 minutes, we need to define intermediate times. Let x represent the time the car takes to travel from the initial position (t_1) to the point where it's directly aligned with the tower's base. Using right triangles, and applying similar trigonometric relationships for the 45° and 60° ...

Text-based analysis

Tedious and time-consuming



(a) Visualization by t-SNE.



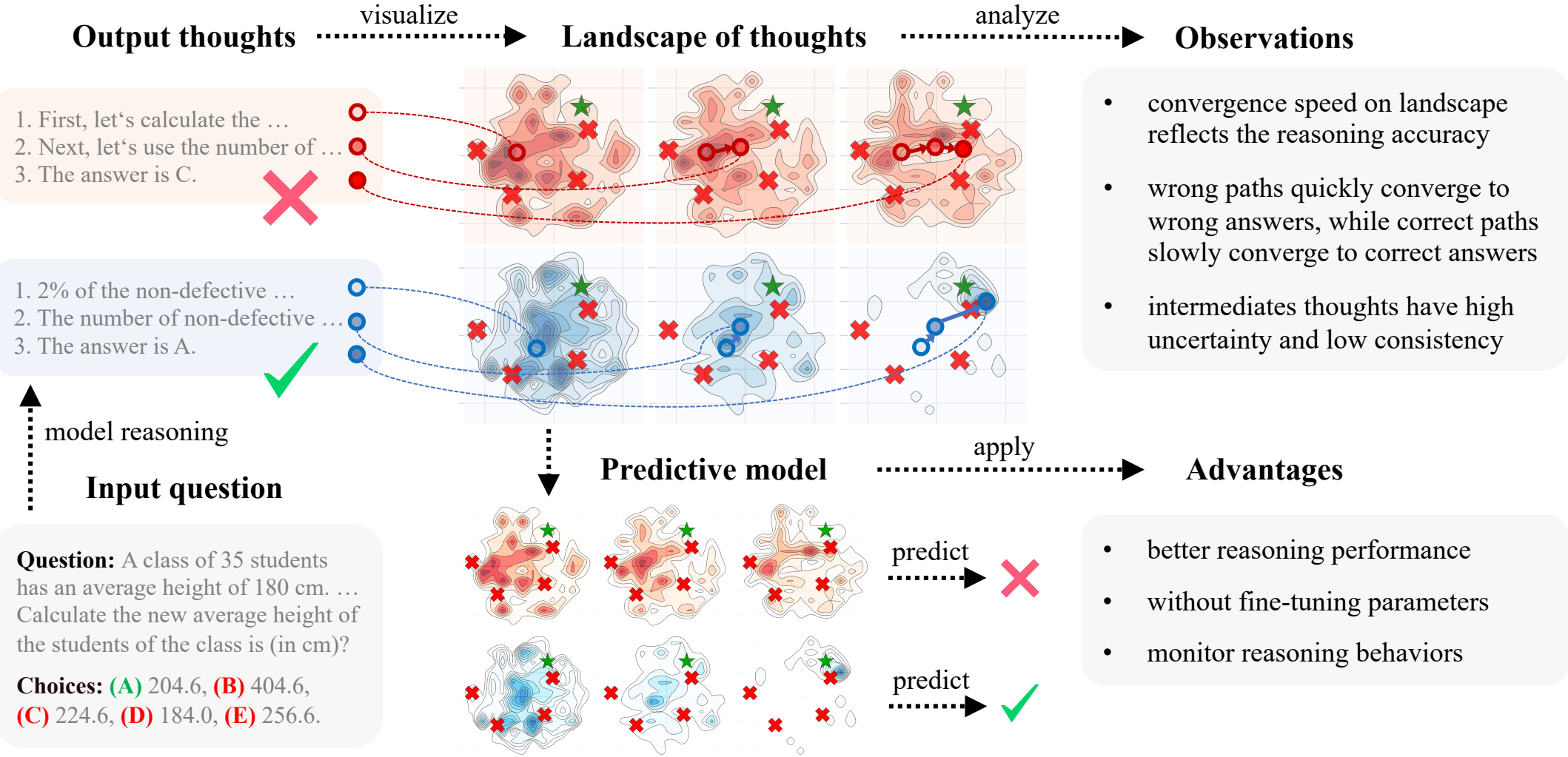
(b) Visualization by Sammon mapping.

Figure 2: Visualizations of 6,000 handwritten digits from the MNIST data set.

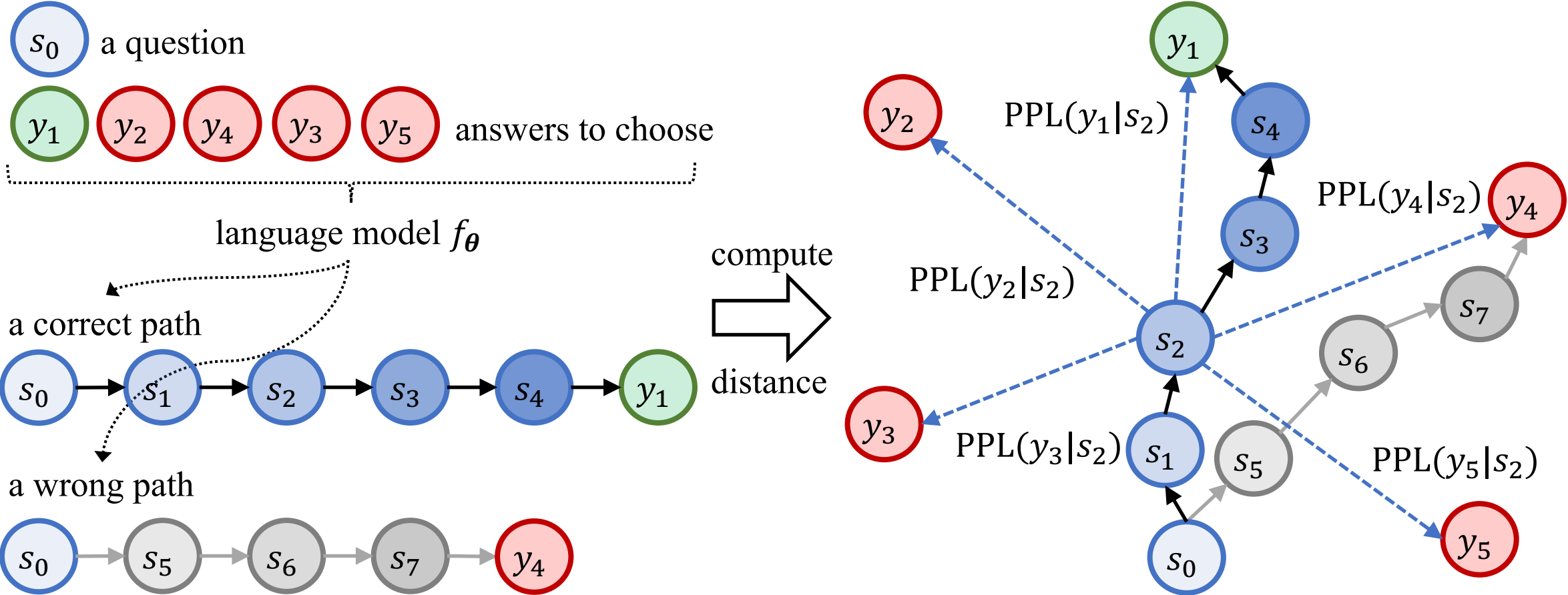
Visualization analysis

More easy and intuitive

Method: How to visualize?



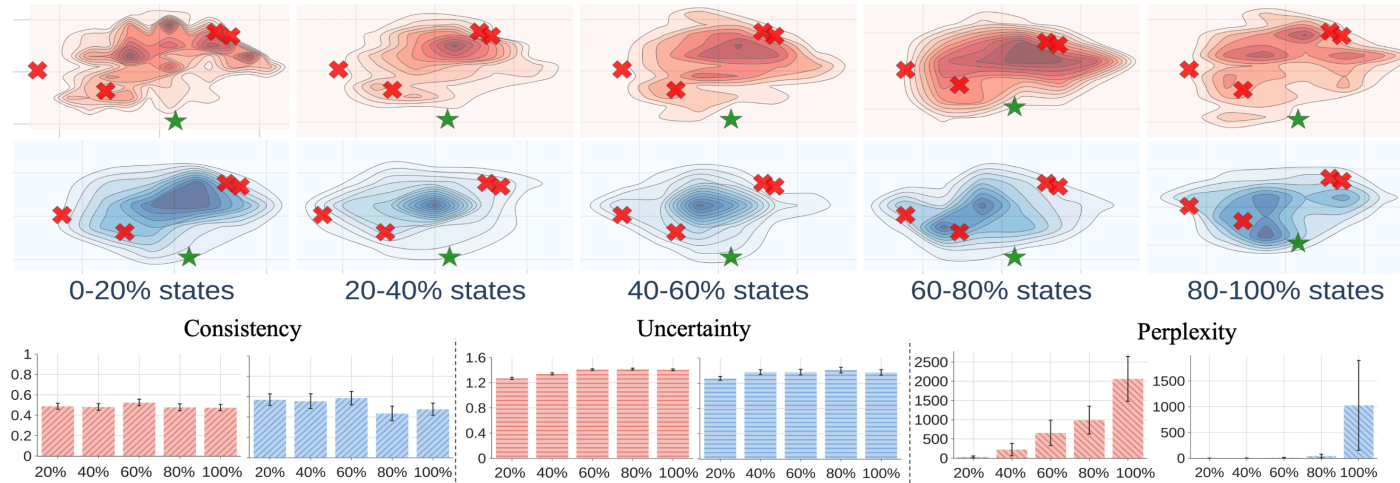
Method: How to visualize?



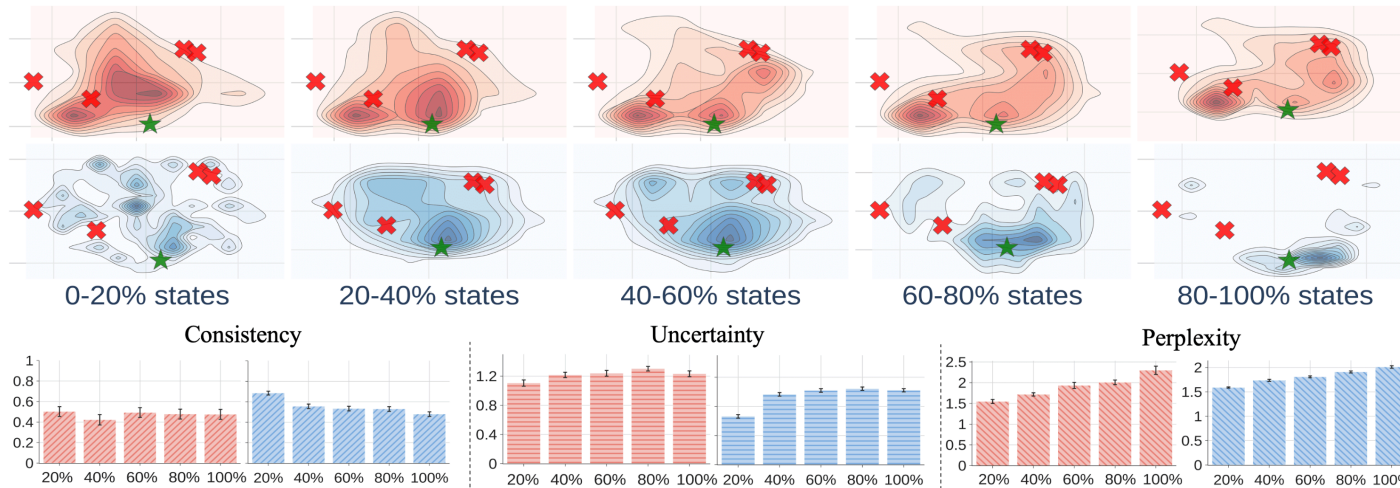
Original thoughts (texts) \rightarrow Feature of distance (values) \rightarrow Low-dimensional representation (values)

Observation: What to visualize?

Llama-3.2-1B



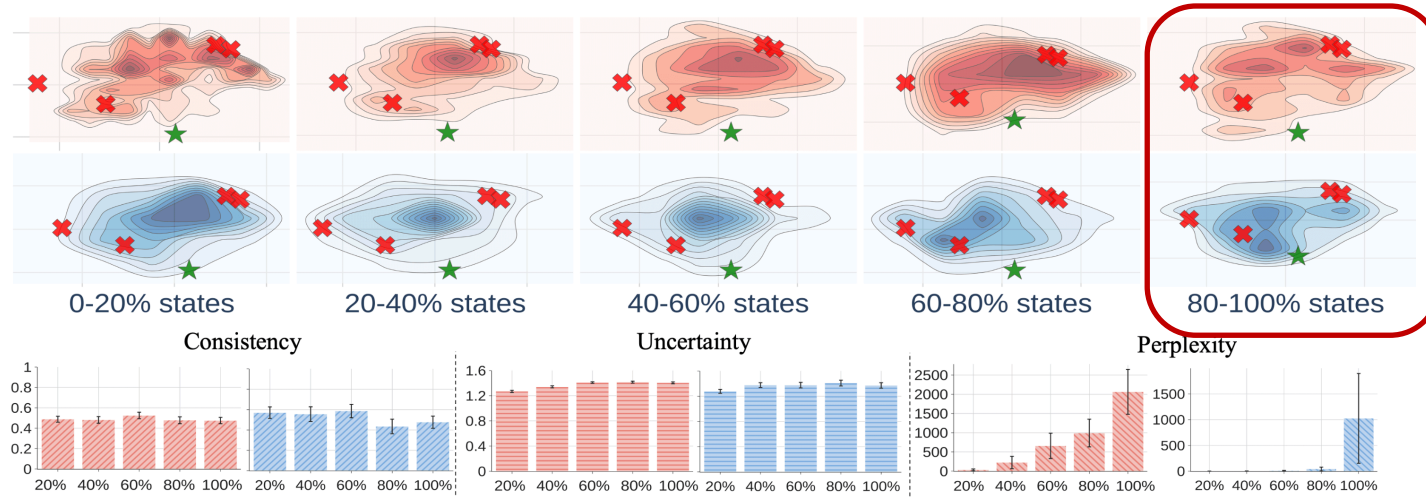
Llama-3.1-70B



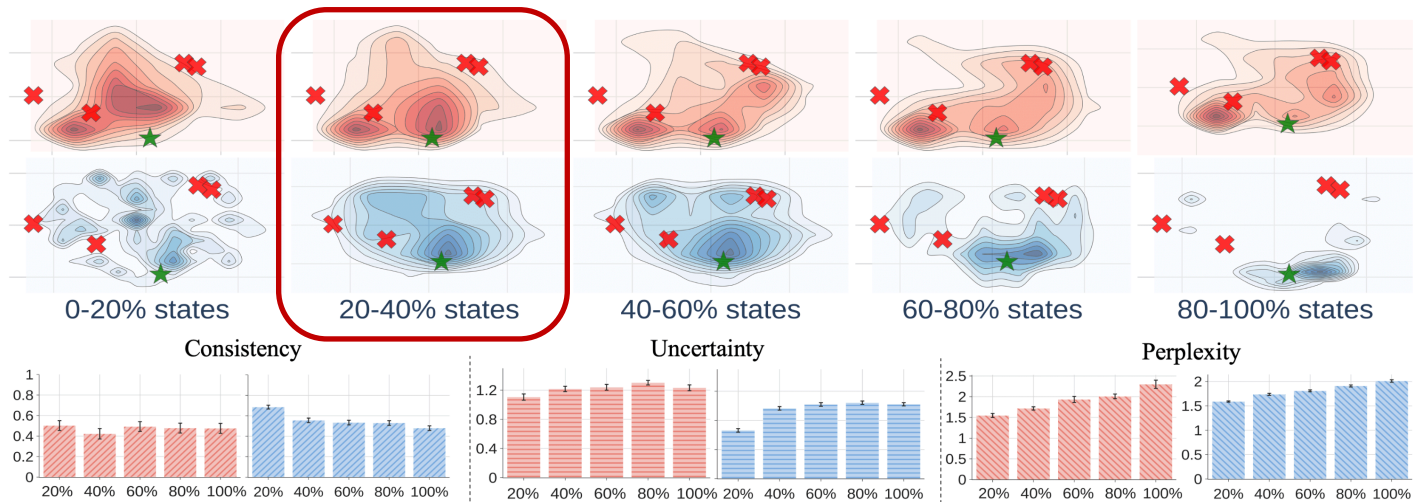
Observation: What to visualize?

(Obs 3.1) The landscape converges faster as the model size increase

Llama-3.2-1B

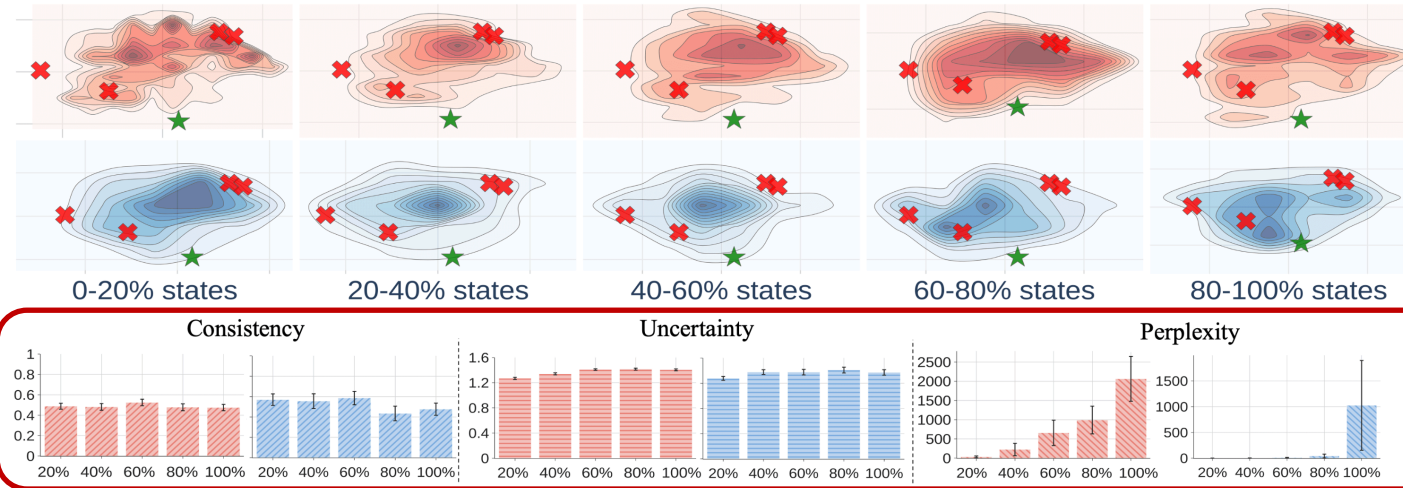


Llama-3.1-70B



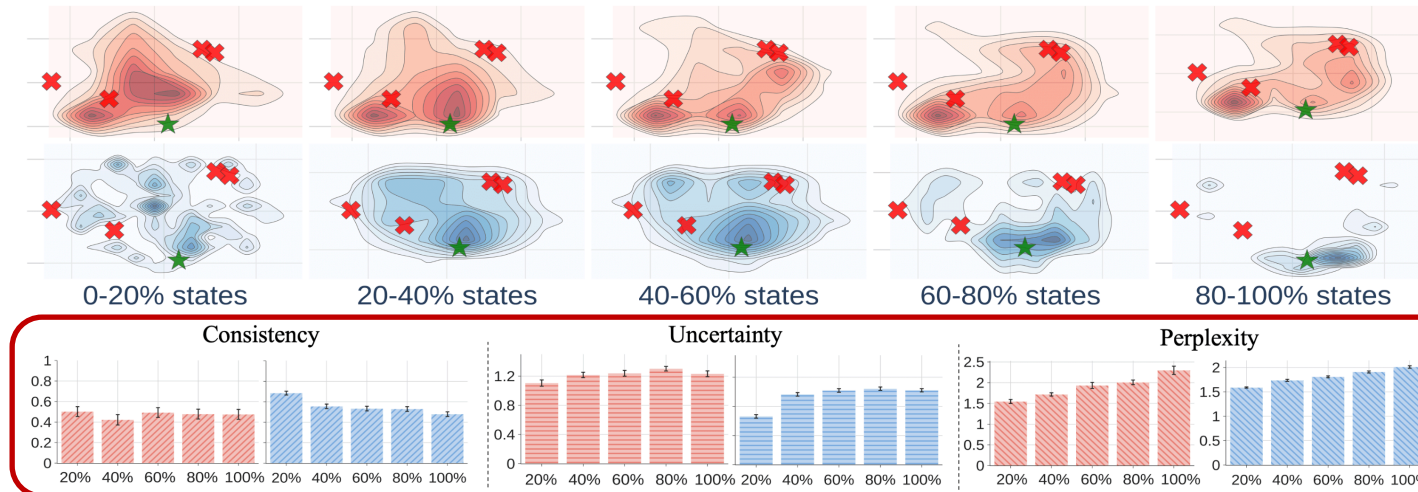
Observation: What to visualize?

Llama-3.2-1B

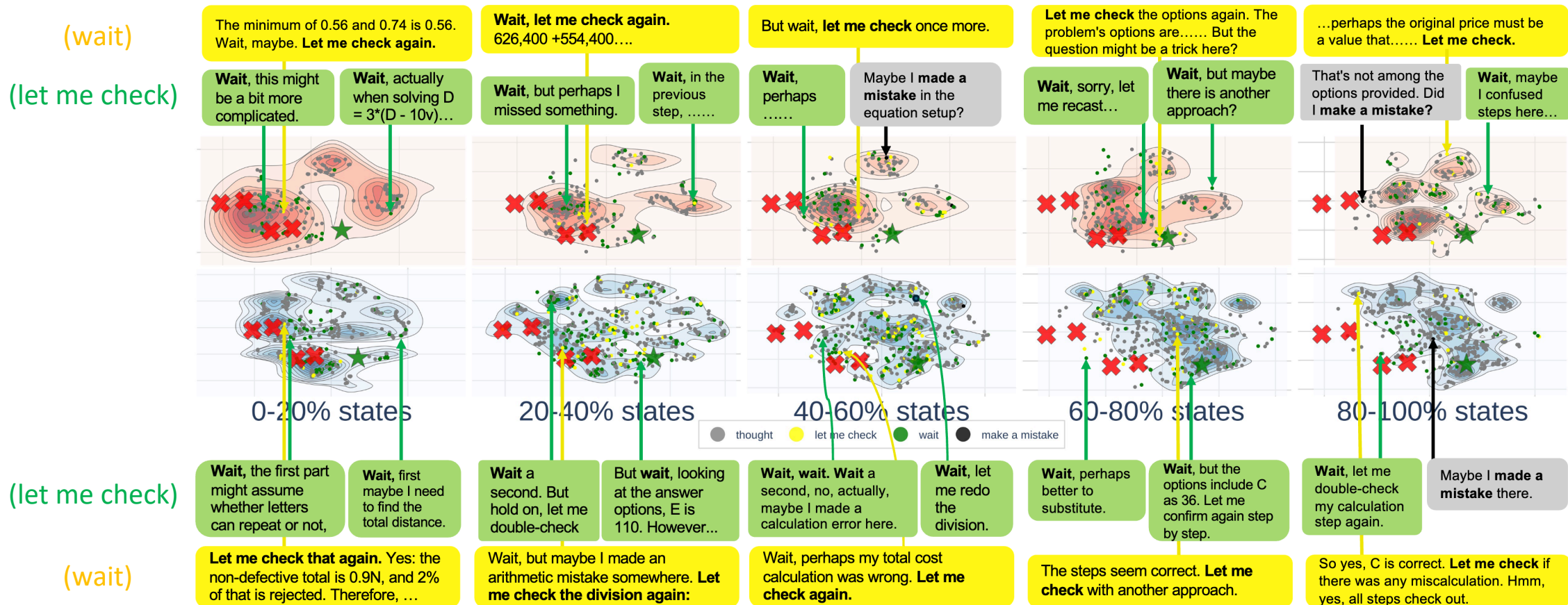


(Obs 3.2) Larger models have higher consistency, lower uncertainty, and lower perplexity

Llama-3.1-70B



Observation: What to visualize?



(Obs 3.3) Reasoning models present more-complex reasoning behaviors in landscapes

Observation: What should be visualized?

Comparison across **language models** (Llama 1B, 3B, 8B, 70B, Reasoning models)

- (Obs 3.1) The landscape converges faster as the model size increase
- (Obs 3.2) Larger models have higher consistency, lower uncertainty, and lower perplexity
- (Obs 3.3) Reasoning models present more-complex reasoning behaviors in landscapes

Comparison across **reasoning tasks** (AQuA, MMLU, StrategyQA, CommonSenseQA)

- (Obs 3.4) Similar reasoning tasks exhibit similar landscapes
- (Obs 3.5) Different reasoning tasks present significantly different patterns in consistency, uncertainty, and perplexity

Comparison across **reasoning algorithms** (CoT, ToT, L2M, MCTS)

- (Obs 3.6) Cross-method comparison: Among correct reasoning trajectories, methods with faster convergence to correct answers achieve higher accuracy
- (Obs 3.7) Within-method comparison: For any single method, incorrect trajectories converge faster to wrong answers than correct trajectories converge to right answers
- (Obs 3.8) Compared to failure trajectories, the intermediate states in correct trajectories have higher consistency w.r.t. the final state