

# Reasoning on a Spectrum: Aligning LLMs to System 1 and System 2 Thinking

Alireza S. Ziabari\*

Nona Ghazizadeh\*

Zhivar Sourati

Farzan Karimi-Malekabadi

Payam Piray

Morteza Dehghani

University of Southern California

{salkhord, nghaziza, souratih, karimima, piray, mdehghan}@usc.edu

## Abstract

Large Language Models (LLMs) exhibit impressive reasoning abilities, yet their reliance on structured step-by-step processing reveals a critical limitation. While human cognition fluidly adapts between intuitive, heuristic (System 1) and analytical, deliberative (System 2) reasoning depending on the context, LLMs lack this dynamic flexibility. This rigidity can lead to brittle and unreliable performance when faced with tasks that deviate from their trained patterns. To address this, we create a dataset of 2,000 samples with valid System 1 and System 2 answers, explicitly align LLMs with these reasoning styles, and evaluate their performance across reasoning benchmarks. Our results reveal an accuracy-efficiency trade-off: System 2-aligned models excel in arithmetic and symbolic reasoning, while System 1-aligned models perform better in commonsense tasks. A mechanistic analysis of model responses shows that System 1 models employ more definitive answers, whereas System 2 models demonstrate greater uncertainty. Interpolating between these extremes produces a monotonic transition in reasoning accuracy, preserving coherence. This work challenges the assumption that step-by-step reasoning is always optimal and highlights the need for adapting reasoning strategies based on task demands.<sup>1</sup>

## 1 Introduction

LLMs have demonstrated remarkable reasoning capabilities, often achieving near-human or even superhuman performance (Huang and Chang, 2023). These advances have largely been driven by techniques that simulate step-by-step, deliberative reasoning, such as Chain-of-Thought (CoT) prompting and inference-time interventions (Wei et al., 2022b; Wang et al., 2022). Given their success,

\*Equal contribution.

<sup>1</sup>Our data and code are available at [https://github.com/AlirezaZiabari/System\\_1\\_System\\_2\\_Alignment](https://github.com/AlirezaZiabari/System_1_System_2_Alignment)

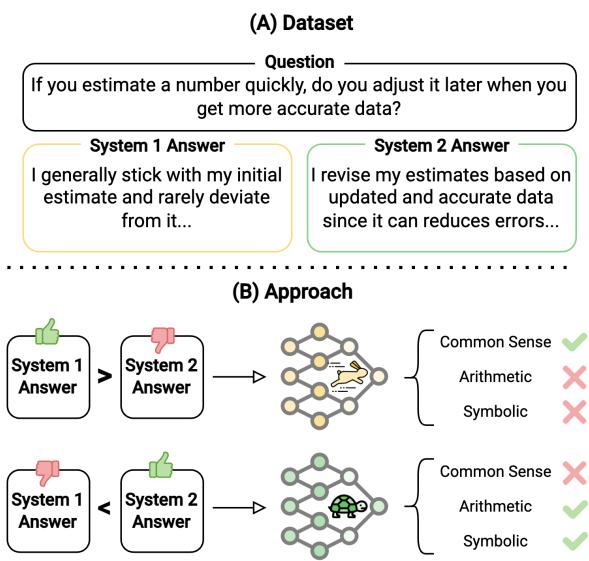


Figure 1: (A) Sample of dataset with System 1 and System 2 answers. (B) Overview of our approach for aligning LLMs with fast and slow thinking, highlighting performance gains across reasoning benchmarks.

such methods are increasingly integrated into LLM training (Chung et al., 2024), reinforcing explicit, structured reasoning regardless of the task necessity. However, LLM reasoning remains brittle, particularly in tasks requiring nuanced judgment (Delétang et al., 2023), logical consistency (Jiang et al., 2024), or adaptability to uncertainty (Mirzadeh et al., 2024). For example, when faced with simple factual queries, they typically generate unnecessarily verbose explanations instead of direct answers (Wang et al., 2023).

This focus on explicit, structured reasoning highlights a key difference between LLMs and human cognition: while LLMs are being pushed towards a single mode of processing, human reasoning is far more nuanced. Rather than a monolithic process, human reasoning emerges from a sophisticated suite of cognitive tools evolved to tackle a *spectrum* of computational problems. This spectrum

of human reasoning encompasses both automatic and reflective processes, a key insight recognized across diverse fields from behavioral economics to psychology and neuroscience (Daw et al., 2005; Dolan and Dayan, 2013; Balleine and Dickinson, 1998). On one end lie computationally *light* problems demanding rapid, intuitive judgments (e.g. instinctively dodging a speeding car), handled by the reflexive “System 1.” On the other end are *heavy* problems requiring deliberate, step-by-step analysis, managed by the reflective “System 2” (Kahneman, 2011; Stanovich and West, 2000). This dual-process system allows us to dynamically shift between modes depending on the task, balancing speed and accuracy (Evans and Stanovich, 2013).

While some studies explore whether LLMs exhibit System 1 and System 2 behaviors (Hagendorff et al., 2023; Pan et al., 2024) or attempt to create hybrid models (Yang et al., 2024; Deng et al., 2024), virtually all prior work implicitly assumes that structured, deliberative reasoning is universally superior. Even research suggesting LLMs’ capacity for both reasoning modes (Wang and Zhou, 2024) largely overlooks the crucial question of when each mode is indeed advantageous. The assumption that a single “best” reasoning strategy can apply across all contexts is a fundamental simplification that limits current approaches in LLM development. This assumption prevents LLMs from achieving true cognitive flexibility, hindering their ability to adapt their reasoning processes to diverse situations.

To address this gap, we propose explicitly aligning LLMs with System 1 and System 2 reasoning and evaluating their reasoning capabilities. Our approach involves designing an experimental setup where both thinking styles can produce valid answers but follow distinct paths, one leveraging intuitive heuristics, and the other prioritizing deliberate, step-by-step reasoning. By systematically assessing the trade-offs between accuracy and efficiency, we provide insights into when intuitive heuristics or structured deliberation is most effective.

Specifically, as demonstrated in Figure 1, we first construct a dataset of 2,000 reasoning tasks, where each problem has both a fast, heuristic-driven (System 1) response and a deliberative, structured (System 2) response, grounded in 10 different cognitive heuristics (Tversky and Kahneman, 1974). We then explicitly align LLMs with either System 1 or System 2 type responses and evaluate these models on diverse reasoning tasks. Our findings reveal a structured accuracy-efficiency trade-off: System 2-

aligned models consistently outperform pre-trained and CoT prompt baselines in arithmetic and symbolic reasoning, demonstrating superior multi-step inference, but generating more extended token-intensive responses. Conversely, System 1-aligned models generate more succinct responses and excel at commonsense reasoning, where heuristic shortcuts are effective. Importantly, unlike CoT models, which always engage in structured reasoning regardless of necessity, our models provide an explicit way to study when different reasoning styles are beneficial, mirroring the well-known efficiency-accuracy trade-off in human cognition (Keramati et al., 2011; Mattar and Daw, 2018). By framing LLM reasoning as a structured and adaptable process, rather than simply an ability to achieve higher benchmark scores, this work highlights the importance of selecting the right reasoning strategy for a given task. This perspective not only aligns LLM reasoning more closely with human cognition but also paves the way for more flexible, efficient, and robust reasoning systems, setting a foundation for future advancements in LLM reasoning.

## 2 Related Work

### 2.1 Reasoning in LLMs

Driven by extensive research highlighting the strengths and weaknesses of LLM reasoning abilities (e.g., Huang and Chang, 2022; Mondorf and Plank, 2024; Valmeekam et al., 2022; Parmar et al., 2024; Sourati et al., 2024), recent efforts to enhance these capabilities have largely focused on prompting techniques (Brown et al., 2020), ranging from zero-shot prompting with explicit instructions (Kojima et al., 2022; Wang et al., 2023; Zhou et al., 2024b) to few-shot prompting with step-by-step examples (Wei et al., 2022b). Wang and Zhou (2024) take CoT prompting even one step further and demonstrate CoT reasoning paths can be elicited from pre-trained LLMs by simply altering the decoding process without the use of a specific prompt. Related approaches, such as self-consistency decoding (Wang et al., 2022), explore how diverse reasoning paths can enhance robustness, aligning with deliberative aspects of System 2 reasoning. Tree of Thought (ToT; Yao et al., 2024) generalizes over CoT and allows LMs to perform deliberate decision making by considering multiple different reasoning paths and self-evaluating choices to decide the next course of action, as well as looking ahead or backtracking when necessary to make a global

choice. Another alternative way of increasing the reasoning abilities of LLMs is through instruction tuning on a substantial amount of CoT reasoning data (Chung et al., 2024; Huang et al., 2022) or distillation (Magister et al., 2022). By training LLMs on a large-scale CoT dataset, models can internalize step-by-step reasoning, potentially enhancing their performance across diverse benchmarks without relying solely on prompting techniques.

## 2.2 Dual-Process Theory in NLP

Dual-process theories, widely studied in psychology, distinguish between fast, intuitive reasoning (System 1) and slow, deliberate reasoning (System 2). While these theories have long explained the spectrum of human reasoning, their application in NLP remains underexplored. Existing research falls into two main categories: (1) analyzing LLMs’ reasoning through the lens of dual-process theory, identifying similarities and differences between LLMs and human reasoning, and (2) developing models that explicitly integrate dual-process mechanisms to enhance LLM reasoning and leverage the benefits of both systems.

**Analyzing LLMs’ reasoning through dual-process theory.** Researchers have investigated whether LLMs exhibit reasoning behaviors aligned with System 1 and System 2, particularly in terms of cognitive human-like errors and biases (Hagendorff et al., 2023; Booch et al., 2021; Pan et al., 2024; Echterhoff et al., 2024; Zeng et al., 2024). Hagendorff et al. (2023) examine cognitive heuristics in LLMs, showing that newer models exhibit fewer errors characteristic of System 1 thinking. Booch et al. (2021) discuss fundamental questions regarding the role of dual-process theory in machine learning but leaves practical implementation as an open problem. Most of these studies evaluate LLMs on benchmarks where System 2 reasoning is assumed to be superior, portraying intuitive responses as erroneous, even though such rapid, heuristic-driven judgments are often crucial for efficient and effective reasoning in real-world scenarios. In contrast, by analyzing the reasoning behavior of models aligned with System 1 and System 2 reasoning—using a carefully curated dataset where both response types are valid—we offer a more nuanced understanding of how this alignment influences broader model behavior.

**Incorporating dual-process theory in NLP models.** Several studies have integrated dual-process-

inspired reasoning into LLMs. Some works combine intuitive (fast) and deliberate (slow) components to improve reasoning (He et al., 2024; Liu et al., 2022; Hua and Zhang, 2022; Pan et al., 2024), while others optimize reasoning efficiency by distilling System 2 insights into System 1 models (Yang et al., 2024; Deng et al., 2024; Yu et al., 2024). Additionally, research has leveraged System 2 reasoning to mitigate biases associated with System 1 heuristics, improving fairness and robustness (Furniturewala et al., 2024; Kamruzzaman and Kim, 2024; Weston and Sukhbaatar, 2023). While prior work largely frames System 2 reasoning as superior or explicitly builds dual-process components within models, our approach investigates the implicit effects of aligning LLMs toward System 1 or System 2 responses. By analyzing how these heuristics influence general reasoning capabilities, we address a gap in the literature and provide new insights into the broader cognitive behaviors of LLMs that have implications for how unseen properties of data that LLMs are trained on can affect their capabilities.

## 3 Method

### 3.1 Aligning to System 1 & System 2 Thinking

We formalize the task of modeling fast or slow thinking as an alignment problem. This approach is effective for two key reasons. First, our aim is not to introduce new knowledge or instructions but rather to shape the model’s reasoning process based on existing capabilities. Second, previous research has shown that prompt engineering can guide LLMs toward System 2 reasoning (Wei et al., 2022a) or System 1 reasoning (Zhou et al., 2024a). This suggests that LLMs already have both reasoning abilities. Therefore, instead of creating new reasoning pathways, our goal is to guide the model to favor either System 1 or System 2 thinking by distinguishing between them.

Within an alignment framework, one response is designated as the preferred (winner) answer, while the other is demoted (loser). To encourage System 1 thinking, we align the model by treating the intuitive response as the winner and the deliberative response as the loser. Conversely, to induce System 2 thinking, we invert this preference, treating the analytical response as the winner and the intuitive response as the loser. This structured preference framework allows the model to internalize the distinction between intuitive and analytical rea-

Category	Question	System 1 Answer	System 2 Answer
Anchoring Bias	When booking travel, do you find it hard to move away from the first deal you see?	I often find myself comparing everything to that initial deal. Once I've encountered a specific offer or price, it becomes the standard by which I measure all others. Even when new deals come along, my mind instinctively revisits that first one, guiding my judgment. It's an automatic comparison that influences how I evaluate options moving forward.	To book travel effectively, I avoid comparisons only against the initial option. I understand that the first deal isn't always the best ensures a broader perspective, more observations increases the chances of finding a more favorable travel plan. Exploring more options ultimately leads to better decision-making.
Halo Effect Bias	Would you say fasting diet is effective?	This diet is obviously effective. just look at the people who follow it! Celebrities and influencers swear by it, and they're in great shape. When so many successful, healthy people use it, that's proof it works. Plus, it's super popular, which wouldn't happen if it didn't give great results. If you want to see real change, this is the way to go!	The effectiveness of a diet depends on scientific evidence, individual health factors, and long-term sustainability rather than its popularity or association with celebrities. A diet may lead to short-term weight loss but could lack essential nutrients or be difficult to maintain. Evaluating its impact on metabolism, overall health, and adherence is crucial..
Over Confidence Bias	Do you believe you will ace the interview?	Yes, I excel in interviews. I communicate clearly, stay confident under pressure, and listen attentively to questions. My ability to understand the interviewer's needs and align my responses accordingly enhances my effectiveness. I maintain engaging body language and make genuine connections, making a lasting impression. I prepare thoroughly, anticipate potential questions, and rehearse answers, ensuring I approach interviews with a calm, composed demeanor, making me a strong candidate.	While confidence in interpersonal skills is beneficial, thorough interview preparation is essential for success. It allows for anticipation of potential questions and crafting informed responses, showcasing an understanding of the company and role. Researching company culture enables candidates to align their answers with organizational values, enhancing their chances of making a positive impression. Solely relying on confidence can lead to unpreparedness, especially for technical inquiries, reducing the effectiveness of skill articulation.

Table 1: A sample of the curated dataset, illustrating the distinction between System 1 and System 2 thinking. Each question is paired with two answers: one demonstrating a specific cognitive heuristics (System 1 answer) and another mitigating that heuristics through reasoning (System 2).

soning processes without requiring any changes to its underlying knowledge. In the next section, we describe the dataset creation process that enables alignment training.

### 3.2 Dataset of System 1 & System 2 Thinking

Our dataset consists of 2,000 questions designed to elicit two distinct response styles in English: one intuitive and rapid, reflecting cognitive shortcuts (System 1), and the other deliberate and analytical (System 2). This dual structure allows us to study the distinct mechanisms underlying System 1 and System 2 reasoning (Kahneman, 2011; Stanovich and West, 2000; Evans and Stanovich, 2013). To create this dataset, we use cognitive heuristics that offer a practical way to distinguish between System 1 and System 2 reasoning, where both produce valid but distinct responses. A sample of the curated dataset is shown in Table 1. These questions span 10 different cognitive heuristics (Appendix A). There are four stages for dataset creation.

**Cognitive heuristics examples.** To create an initial set of examples, an expert selected 10 categories of heuristics and biases from Kahneman (2011) and generated one example question for each category with both System 1 and System 2 answers. These examples are in Appendix B.

**Data expansion using an LLM.** We then expanded the dataset with GPT-4o (Hurst et al., 2024) with one-shot prompting. For each cognitive heuristic, we provided its definition (Kahneman, 2011), a description of how System 1 and System 2 thinking would approach the question (Kahneman, 2011),

and an expert-generated example illustrating the heuristic. More details about the prompt for data expansion are outlined in Appendix C

**Human validation and refinement.** To ensure the dataset accurately reflected our definitions of fast and slow thinking, as well as the targeted cognitive heuristics, the expert manually reviewed and modified around 20% of the responses.

**Length adjustment.** As a result of the data expansion process, System 2 answers tended to be longer and more detailed, reflecting their step-by-step reasoning approach, whereas System 1 answers were shorter and more direct, relying on heuristics. The data expansion process yielded a significant difference in answer length between the two reasoning styles, with System 2 answers being substantially longer. This difference in length was verified using Welch's  $t$ -test ( $t(2090.1) = -184.74$ ,  $p < .001$ ,  $d = -5.84$ ). This test was chosen due to unequal variances between the two answer types confirmed by Levene's test ( $F(1, 3998) = 2487.9$ ,  $p < .001$ ).

Recently, Singhal et al. (2023) highlight a strong correlation between output length and perceived quality, with longer outputs often being considered preferable. To address this potential problem, we used GPT-4o with zero-shot prompting to adjust the lengths of System 1 and System 2 answers, ensuring comparability without changing their content. This adjustment was applied only when there was a significant length disparity. More details about the prompt and length disparity threshold are described in Appendix E. By reducing the length disparity,

we minimized any preference for System 2 answers arising from their longer responses. After adjustment, System 1 answers had an average length of 82.19 tokens, while System 2 answers averaged 83.93 tokens. A two one-sided t-test (TOST) confirmed the equivalence of these post-adjustment lengths across various token counts as equivalence margins,<sup>2</sup> indicating that the adjustment effectively eliminated significant length differences between the two response types.

## 4 Experiments

### 4.1 Alignment Algorithm

To implement the alignment strategy for System 1 and System 2 reasoning, we utilize two prominent preference optimization methods:

Direct Preference Optimization (DPO; Rafailov et al., 2024) is an offline alignment method that fine-tunes LLMs by comparing the preferred and disfavored outputs of a model against a reference model, optimizing preferences without requiring a separate reward model. As a prominent method in preference optimization, DPO has gained traction for its stability and efficiency, making it a widely adopted alternative to Reinforcement Learning from Human Feedback (RLHF; Ouyang et al., 2022).

Simple Preference Optimization (SimPO; Meng et al., 2024) builds on the principles of DPO but introduces a reference-free approach to preference optimization. Instead of requiring a separate reference model, SimPO aligns responses by directly optimizing preference signals within the model itself. This makes it computationally more efficient and removes the dependency on an external reference model, offering a streamlined alternative for aligning LLMs to a specific preference.

### 4.2 Benchmarks

Building on previous research in LLM reasoning (Wei et al., 2022b; Kojima et al., 2022; Kong et al., 2024), we evaluate our System 1 and System 2 models using 10 reasoning benchmarks across three different categories: (1) arithmetic reasoning, which includes MultiArith (Roy and Roth, 2015), GSM8K (Cobbe et al., 2021), AddSub (Hosseini et al., 2014), AQUA-RAT (Ling et al., 2017), SingleEq (Koncel-Kedziorski et al., 2015), and SVAMP (Patel et al., 2021); (2) commonsense reasoning, in-

<sup>2</sup> $\pm 3$  tokens,  $t(3870.30) = 85.82, p < .001$ ;  $\pm 5$  tokens,  $t(3870.30) = 149.07, p < .001$ ;  $\pm 7$  tokens,  $t(3870.30) = 212.31, p < .001$ ; and 5% of the mean token count ( $\pm 4.15$  tokens),  $t(3870.30) = 122.29, p < .001$

cluding CSQA (Talmor et al., 2019) and StrategyQA (Geva et al., 2021); (3) symbolic reasoning, covering Last Letter Concatenation and Coin Flip (Wei et al., 2022b). More details about the benchmarks are in Appendix D.

Following Kong et al. (2024), our evaluation follows a two-stage process. In the first stage, we present the benchmark questions to the model and record its responses. In the second stage, we prompt the model again, this time providing the original question, its initial response, and benchmark-specific instructions to ensure the output is formatted as required. The instructions for each benchmark are detailed in Appendix F.

### 4.3 Experiment Setup & Details

We use Llama-3-8B-Instruct (AI@Meta, 2024) and Mistral-7B-Instruct-v0.1 (Jiang et al., 2023) as the SFT models in the alignment process. Following Kojima et al. (2023), we compare the performance of these aligned models against their pre-trained counterparts under zero-shot and zero-shot CoT prompting (additional details in Appendix G).

To analyze the model’s behavior along the System 1-System 2 reasoning spectrum, we train seven intermediate models, where the winner responses are mixed at predefined ratios: 87.5%-12.5%, 75%-25%, 62.5%-37.5%, 50%-50%, 37.5%-62.5%, 25%-75%, and 12.5%-87.5%, between System 1 and System 2. This structured variation allows us to systematically examine the transition between System 1 and System 2 reasoning.

## 5 Results

### 5.1 Comparing Reasoning Benchmarks

Table 2 presents a comparison of exact matching accuracy across 10 reasoning benchmarks for two LLMs, Llama and Mistral. Specifically, we compare the base models with the System 1 and System 2 variants created using the alignment algorithms described in Section 4.1. We also include results for CoT prompting for reference. Our findings reveal distinct performance trends for the System 1 and System 2 models, highlighting their respective strengths in different reasoning paradigms.

Across all arithmetic benchmarks (MultiArith, GSM8K, AddSub, AQuA, and SingleEq), System 2 models consistently outperformed both the base model and their System 1 counterpart, evident for both Llama and Mistral. This improvement is most significant in AddSub and SingleEq benchmarks.

		Arithmetic					Symbolic		Common Sense		
		MultiArith	GSM8K	AddSub	AQuA	SingleEq	SVAMP	Coin	Letter	CSQA	Strategy
System 2	DPO	98.67 (+1.0)	79.37 (+0.88)	89.87 (+7.4)	49.21 (+0.39)	94.37 (+3.65)	85.4 (+4.9)	93.8 (-0.4)	86.2 (+2.2)	71.42 (0)	60.87 (-6.68)
	SIMPO	97.83 (+0.16)	79.38 (+0.89)	90.13 (+7.66)	54.72 (+6.78)	94.49 (+3.77)	81.7 (+1.2)	94.4 (+0.2)	84.8 (+0.8)	69.62 (-1.8)	67.38 (-0.17)
	Llama-3	97.67	78.49	82.47	48.82	90.72	80.5	94.2	84	71.42	67.55
System 1	Llama-3-CoT	97.83	78.54	82.03	49.21	88.19	80.9	94.8	84.2	71.58	67.38
	DPO	98.5 (+0.83)	77.01 (-1.48)	80.76 (-1.71)	46.46 (-2.36)	77.24 (-13.48)	78 (-2.5)	93.4 (-0.8)	83.8 (-0.2)	72.81 (+1.39)	68.21 (+0.66)
	SIMPO	97.5 (-0.17)	77.79 (-0.7)	80.51 (-1.96)	48.03 (-0.79)	87.4 (-3.32)	79.3 (-1.2)	90 (-4.2)	83.8 (-0.2)	72.32 (+0.9)	67.73 (+0.18)
System 2	DPO	78.83 (+1.16)	56.45 (+1.47)	81.27 (+6.79)	32.68 (+1.19)	84.84 (+0.98)	69.1 (+3.4)	41 (-2.2)	8.6 (+8)	62.82 (-3.44)	56.81 (-8.6)
	SIMPO	78.3 (+0.63)	55.42 (+0.53)	82.28 (+7.8)	34.25 (+2.76)	86.81 (+2.95)	68.5 (+2.8)	45.4 (+2.2)	7.8 (+6.2)	64.78 (-1.48)	63.75 (-1.66)
	Mistral	77.67	54.89	79.75	31.49	83.86	66.26	43.2	1.6	67.65	65.41
System 1	Mistral-CoT	78.3	54.96	80.25	33.07	83.66	67.8	43.8	1.6	66.18	65.49
	DPO	77.5 (-0.17)	51.4 (-3.49)	79.49 (-0.26)	29.53 (-1.96)	83.07 (-0.79)	67.4 (-0.2)	40.4 (-2.8)	0 (-1.6)	67.4 (+1.14)	65.49 (+0.08)
	SIMPO	77 (-0.67)	53.61 (-1.28)	78.73 (-1.02)	31.1 (-0.39)	83.67 (-0.19)	67.3 (-0.3)	43 (-0.2)	0 (-1.6)	67.32 (+1.06)	65.51 (+0.1)

Table 2: Accuracy comparison of our System 1 and System 2-aligned models (DPO, SIMPO) against pre-trained and zero-shot CoT baselines (Llama, Mistral) across benchmarks. Each cell shows accuracy, with parentheses indicating the difference from the baseline. Color intensity reflects the magnitude of deviation, with darker shades representing larger differences.

Similarly, System 2 models outperformed System 1 models in nearly all symbolic reasoning tasks. These tasks require pattern recognition and logical structuring, further validating the idea that deliberative, slow-thinking models enhance performance in structured reasoning.

Conversely, System 1 models excelled at commonsense reasoning tasks, particularly on the CSQA and Strategy benchmarks. These tasks, which rely on intuitive judgment, and heuristic decision-making, clearly play to System 1’s strengths. Both Llama and Mistral saw their System 1 variants outperform not only their System 2 counterparts but also their respective base and CoT models in these domains.

When comparing Llama and Mistral, Llama models generally achieved higher accuracy across all benchmarks. This suggests that Llama may have stronger foundational reasoning capabilities, which are further enhanced by the System 2 and System 1 alignment. Moreover, pre-trained models with CoT prompt generally showed improvements over base models, reinforcing the effectiveness of explicit step-by-step reasoning. However, since pre-trained models have already been trained on CoT data, the improvements from explicit prompting are not significant. This suggests that step-by-step thinking has been internalized within these pre-trained models, reducing the necessity for additional CoT prompts. Based on this observation, we focus solely on Llama models in the following

experiments.

In summary, our results showcase that System 2 models excel in structured, multi-step reasoning tasks such as arithmetic and symbolic reasoning, while System 1 models are particularly effective in intuitive and commonsense reasoning tasks. These findings highlight the significant potential of dual-process alignment for boosting LLM performance across a diverse range of reasoning paradigms.

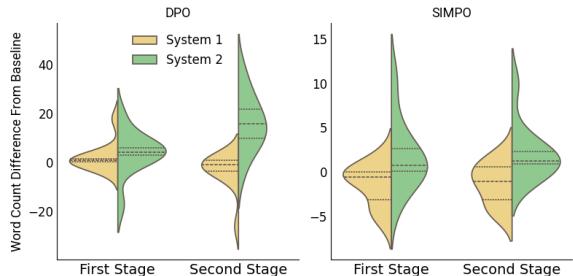


Figure 2: Comparison of the token difference in System 1 and System 2 answers relative to the Llama3 model across both response stages and for both alignment algorithms, DPO and SimPO.

## 5.2 Model Response Analysis

As described in Section 4.2, we use a two-stage prompting to generate the final responses from our models. Figure 2 shows the difference in token counts for System 1 and System 2 answers, relative to the Llama model, across both stages for the two alignment algorithms, DPO and SimPO. Although both System 1 and System 2 models are aligned

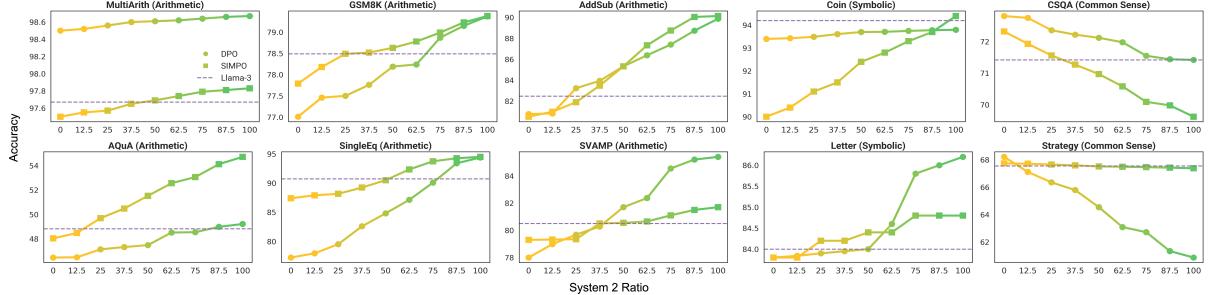


Figure 3: Accuracy across different benchmarks as reasoning shifts from System 1 to System 2.

based on the equal-length response as mentioned in Section 3.2, the average length of tokens in System 1 and System 2 answers differs in the second stage for both DPO,  $t(8836) = 57.14, p < .001$ , and SimPO,  $t(8586) = 9.833, p < .001$ . This suggests that the System 2-aligned model does not rely solely on its initial response and attempts to elaborate, resulting in more accurate answers when this additional rigor is beneficial.

### 5.3 Moving from Fast to Slow Thinking

In the previous study, we considered System 1 and System 2 as endpoints on a spectrum. Paralleling psychological approaches (Daw et al., 2011; Piray and Daw, 2021), we explored the space between these extremes by creating interpolated models by blending System 2 and System 1 answers at varying ratios. Figure 3 demonstrates a consistent, monotonic increase in accuracy across all benchmarks ( $r^2 > 0.9, p < 0.001$ ). Critically, there are no sudden drops or fluctuations in performance when transitioning between reasoning styles. This stability indicates that the shift from System 1 to System 2 reasoning is gradual and predictable, without any unexpected anomalies. This observation reinforces the idea that LLMs can be strategically guided toward different reasoning styles without sacrificing the coherence of their responses.

While arithmetic and symbolic reasoning tasks exhibit a steady increase in accuracy moving toward System 2 thinking, commonsense tasks show the opposite trend, with accuracy increasing as models rely more on System 1 reasoning. This trade-off highlights that both reasoning styles offer unique advantages, with System 2 excelling in structured, multi-step problem-solving and System 1 providing efficient, adaptable responses in intuitive scenarios. These findings straighten the importance of task-dependent reasoning strategies that leverage the strengths of both System 1 and

System 2 thinking.

### 5.4 Reasoning & Uncertainty

Beyond evaluating model reasoning abilities based on reasoning benchmarks, we examine how uncertainty manifests in System 1 and System 2 responses. A key insight from psychology and neuroscience is that System 1 operates on confident heuristics, providing quick, intuitive judgments, while System 2 engages in more deliberate, analytical thought, accurately assessing the uncertainty associated with its conclusions (Daw et al., 2005; Lee et al., 2014; Keramati et al., 2011; Xu, 2021). To examine uncertainty and confidence, we consider three different characteristics: 1) token-level uncertainty; 2) the presence of hedge words (Lakoff, 1973; Ott, 2018) such as “*might*” and “*possibly*”<sup>3</sup> in model output; and 3) definitive commitment to answers in System 1 versus System 2.

Measuring token-level uncertainty through the logits of generated tokens, Figure 4, Plot A shows that System 2-aligned models consistently generate tokens with lower confidence compared to System 1 models. This trend holds across arithmetic  $t(4075) = 54.53, p < .001$ , symbolic  $t(999) = 42.53, p < .001$ , and commonsense  $t(3510) = 106.86, p < .001$  benchmarks. Additionally, we analyzed surface-level uncertainty in model reasoning by examining word choices. Figure 4, Plot B shows System 2-aligned models use significantly more hedge words, in arithmetic  $t(4075) = 22.03, p < .001$  and commonsense  $t(3510) = 21.49, p < .001$  when models reiterate their reasoning. While increased uncertainty enhances analytical reasoning, it may hinder tasks requiring rapid, intuitive judgments. To assess early-stage response conclusiveness, we used LLM-as-Judge (Zheng et al., 2023) as detailed in Appendix H. Figure 4, Plot C shows System 1 mod-

<sup>3</sup>Gathered from <https://github.com/words/hedges>.

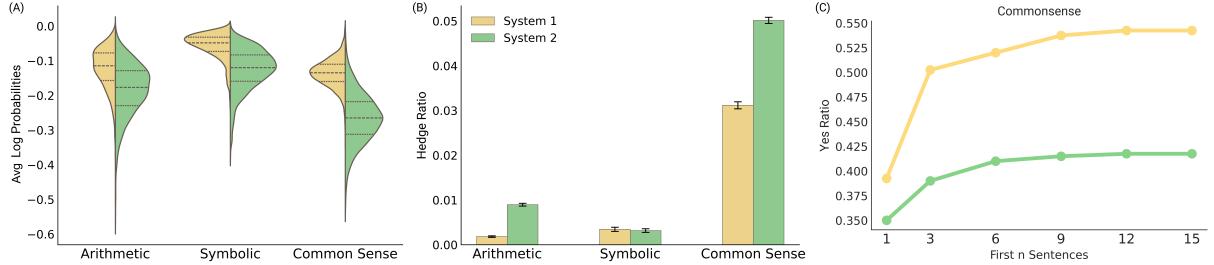


Figure 4: (A) Average log probabilities of models’ reasoning representing their inherent uncertainty in its reasoning; (B) Ratio of hedge words in models’ reasoning representing their surface-level uncertainty; (C) Proportion of definitive answers in the first n sentences.

els provide significantly more definitive answers than System 2 models in commonsense reasoning, *McNemar’s*  $\chi^2(1, 400) = 20.0, p < .001$ , regardless of where in the response the definitive answer is reached (see Appendix H).

This analysis reinforces the idea that different reasoning styles are suited to different tasks. Greater uncertainty in models’ generated reasoning suggests that System 2 models can explore alternative reasoning paths more effectively. This uncertainty is reflected in both their model output probabilities and word choices. System 2 models’ superior performance in arithmetic tasks highlights the benefits of deliberate, effortful processing in tasks that demand exploration and uncertainty. On the other hand, the greater tendency of System 1 models to commit to answers in a more definitive way aligns with their advantage in tasks requiring rapid and intuitive judgments. This behavior is observed exclusively in commonsense reasoning, where quick, decisive responses are advantageous—a trend supported by human studies (Byrd, 2022) and confirmed by our findings in Section 5.1. However, it does not appear in other benchmarks (see Appendix H), suggesting that the activation of a particular reasoning style is context-dependent and influenced by task demands.

## 6 Conclusion

We proposed aligning LLMs to System 1 and System 2 thinking, representing two ends of the spectrum in fast and slow cognitive processes. Our results demonstrate that each of these models exhibits distinct properties: System 2 excels in arithmetic and symbolic reasoning, while System 1 is more effective and accurate in commonsense reasoning (Section 5.1). By introducing intermediate models that interpolate between System 1 and System 2, we observed a monotonic accuracy shift across

all benchmarks, reinforcing the idea that reasoning ability transitions smoothly along this spectrum (Section 5.3). Additionally, System 1 models generate responses with fewer tokens, highlighting its efficiency in decision-making (Section 5.2). Finally, our analysis in Section 5.4 illustrated that System 2 models exhibit greater uncertainty throughout the reasoning process, potentially enabling them to engage in more structured, step-by-step problem-solving. In contrast, System 1 models display higher confidence, allowing them to reach answers faster, which is particularly advantageous for tasks requiring rapid, intuitive judgments.

Beyond these empirical findings, our study aligns with broader principles observed across cognitive science and neuroscience. The observation that System 1 models generate faster responses echoes established theories in human cognition, where intuitive, heuristic-driven thinking allows for rapid decision-making. Similarly, the higher uncertainty exhibited by System 2 models aligns with neuroscience findings that deliberate reasoning involves increased cognitive load and self-monitoring mechanisms. These parallels suggest that LLMs, when properly aligned, can mirror key aspects of human cognition, offering new insights into both artificial and natural intelligence.

This work is a first step toward adaptive reasoning in LLMs, where models can dynamically shift between heuristic and deliberative thinking based on task demands. Furthermore, understanding how to optimally balance speed and accuracy in LLMs can have significant implications for real-world applications, from conversational agents to automated decision-making systems. By grounding our findings in cognitive science, we provide an approach for developing more flexible models that can strategically deploy different reasoning styles on various tasks to maximize efficiency and effectiveness.

## Limitations

Despite the promising advancements of using different thinking styles presented in our approach, several limitations should be acknowledged. First, our curated dataset of 2,000 questions, though designed to capture diverse cognitive heuristics, may not fully represent the entire spectrum of reasoning challenges encountered in real-world tasks. Second, the reliance on prompt engineering and specific alignment algorithms (DPO and SimPO) means that our findings may be sensitive to changes in model architecture or training procedures, potentially limiting generalizability across different LLMs. Third, our analysis of uncertainty—using token-level logits and heuristic word choices—provides a useful proxy but may not comprehensively capture the nuanced aspects of human-like uncertainty. Finally, while our experiments reveal a clear accuracy-efficiency trade-off between intuitive and deliberative reasoning, the extent to which these findings translate to more complex or dynamic decision-making scenarios remains an open question. Future work should explore larger, more diverse datasets and investigate alternative alignment strategies to further validate and extend these results.

## Ethical Statement

This research was conducted with ethical considerations in data collection, model alignment, and societal impact. All cognitive heuristic examples and dataset validation were performed by the authors to guarantee quality but potentially introduce biases.

Aligning LLMs with System 1 and System 2 reasoning raises concerns about model behavior in different contexts. System 1 models may produce overly confident but incorrect answers, while System 2 models, though more deliberate, may slow response times and increase computational costs. Responsible deployment requires balancing these trade-offs to prevent biased or misleading outputs.

## Acknowledgments

This research was supported by DARPA INCAS HR001121C0165. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of DARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute

reprints for governmental purposes notwithstanding any copyright annotation therein.

## References

- Marah Abdin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J Hewett, Mojan Javaheripi, Piero Kauffmann, et al. 2024. Phi-4 technical report. *arXiv preprint arXiv:2412.08905*.
- AI@Meta. 2024. [Llama 3 model card](#).
- Meysam Alizadeh, Maël Kubli, Zeynab Samei, Shirin Dehghani, Juan Diego Bermeo, Maria Krobeynikova, and Fabrizio Gilardi. 2023. Open-source large language models outperform crowd workers and approach chatgpt in text-annotation tasks. *arXiv preprint arXiv:2307.02179*, 42.
- Bernard W Balleine and Anthony Dickinson. 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37(4-5):407–419.
- Grady Booch, Francesco Fabiano, Lior Horesh, Kiran Kate, Jonathan Lenchner, Nick Linck, Andreas Loreggia, Keerthiram Murgesan, Nicholas Mattei, Francesca Rossi, et al. 2021. Thinking fast and slow in ai. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15042–15046.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- Nick Byrd. 2022. Bounded reflectivism and epistemic identity. *Metaphilosophy*, 53(1):53–69.
- Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, et al. 2024. Scaling instruction-finetuned language models. *Journal of Machine Learning Research*, 25(70):1–53.
- Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reichiro Nakano, Christopher Hesse, and John Schulman. 2021. [Training verifiers to solve math word problems](#). *Preprint*, arXiv:2110.14168.
- Nathaniel D Daw, Samuel J Gershman, Ben Seymour, Peter Dayan, and Raymond J Dolan. 2011. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215.
- Nathaniel D Daw, Yael Niv, and Peter Dayan. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, 8(12):1704–1711.

- Grégoire Delétang, Anian Ruoss, Jordi Grau-Moya, Tim Genewein, Li Kevin Wenliang, Elliot Catt, Chris Cundy, Marcus Hutter, Shane Legg, Joel Veness, and Pedro A. Ortega. 2023. Neural networks and the chomsky hierarchy. In *11th International Conference on Learning Representations*.
- Yongxin Deng, Xihe Qiu, Xiaoyu Tan, Chao Qu, Jing Pan, Yuan Cheng, Yinghui Xu, and Wei Chu. 2024. Cognidual framework: Self-training large language models within a dual-system theoretical framework for improving cognitive tasks. *arXiv preprint arXiv:2409.03381*.
- Ray J Dolan and Peter Dayan. 2013. Goals and habits in the brain. *Neuron*, 80(2):312–325.
- Jessica Echterhoff, Yao Liu, Abeer Alessa, Julian McAuley, and Zexue He. 2024. Cognitive bias in decision-making with llms. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 12640–12653.
- Jonathan St BT Evans and Keith E Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3):223–241.
- Shaz Furniturewala, Surgan Jandial, Abhinav Java, Pragyan Banerjee, Simra Shahid, Sumit Bhatia, and Kokil Jaidka. 2024. Thinking fair and slow: On the efficacy of structured prompts for debiasing language models. *arXiv preprint arXiv:2405.10431*.
- Mor Geva, Daniel Khashabi, Elad Segal, Tushar Khot, Dan Roth, and Jonathan Berant. 2021. Did aristotle use a laptop? a question answering benchmark with implicit reasoning strategies. *Transactions of the Association for Computational Linguistics*, 9:346–361.
- Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120.
- Thilo Hagendorff, Sarah Fabi, and Michal Kosinski. 2023. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in chatgpt. *Nature Computational Science*, 3(10):833–838.
- Tao He, Lizi Liao, Yixin Cao, Yuanxing Liu, Ming Liu, Zerui Chen, and Bing Qin. 2024. Planning like human: A dual-process framework for dialogue planning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4768–4791, Bangkok, Thailand. Association for Computational Linguistics.
- Mohammad Javad Hosseini, Hannaneh Hajishirzi, Oren Etzioni, and Nate Kushman. 2014. Learning to solve arithmetic word problems with verb categorization. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 523–533, Doha, Qatar. Association for Computational Linguistics.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- Wenyue Hua and Yongfeng Zhang. 2022. System 1 + system 2 = better world: Neural-symbolic chain of logic reasoning. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 601–612, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Jiaxin Huang, Shixiang Shane Gu, Le Hou, Yuexin Wu, Xuezhi Wang, Hongkun Yu, and Jiawei Han. 2022. Large language models can self-improve. *arXiv preprint arXiv:2210.11610*.
- Jie Huang and Kevin Chen-Chuan Chang. 2022. Towards reasoning in large language models: A survey. *arXiv preprint arXiv:2212.10403*.
- Jie Huang and Kevin Chen-Chuan Chang. 2023. Towards reasoning in large language models: A survey. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1049–1065, Toronto, Canada. Association for Computational Linguistics.
- Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Osztrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.
- Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.
- Bowen Jiang, Yangxinyu Xie, Zhuoqun Hao, Xiaomeng Wang, Tanwi Mallick, Weijie J Su, Camillo Jose Taylor, and Dan Roth. 2024. A peek into token bias: Large language models are not yet genuine reasoners. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4722–4756, Miami, Florida, USA. Association for Computational Linguistics.
- Daniel Kahneman. 2011. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York.
- Mahammed Kamruzzaman and Gene Louis Kim. 2024. Prompting techniques for reducing social bias in llms through system 1 and system 2 cognitive processes. *arXiv preprint arXiv:2404.17218*.
- Mehdi Keramati, Amir Dezfooli, and Payam Piray. 2011. Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS computational biology*, 7(5):e1002055.

- Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213.
- Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2023. [Large language models are zero-shot reasoners](#). Preprint, arXiv:2205.11916.
- Rik Koncel-Kedziorski, Hannaneh Hajishirzi, Ashish Sabharwal, Oren Etzioni, and Siena Dumas Ang. 2015. [Parsing algebraic word problems into equations](#). *Transactions of the Association for Computational Linguistics*, 3:585–597.
- Aobo Kong, Shiwan Zhao, Hao Chen, Qicheng Li, Yong Qin, Ruiqi Sun, Xin Zhou, Enzhi Wang, and Xiaohang Dong. 2024. [Better zero-shot reasoning with role-play prompting](#). Preprint, arXiv:2308.07702.
- George Lakoff. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of philosophical logic*, 2(4):458–508.
- Sang Wan Lee, Shinsuke Shimojo, and John P O’doherity. 2014. Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, 81(3):687–699.
- Wang Ling, Dani Yogatama, Chris Dyer, and Phil Blunsom. 2017. [Program induction by rationale generation: Learning to solve and explain algebraic word problems](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 158–167, Vancouver, Canada. Association for Computational Linguistics.
- Zhixuan Liu, Zihao Wang, Yuan Lin, and Hang Li. 2022. A neural-symbolic approach to natural language understanding. arXiv preprint arXiv:2203.10557.
- Lucie Charlotte Magister, Jonathan Mallinson, Jakub Adamek, Eric Malmi, and Aliaksei Severyn. 2022. Teaching small language models to reason. arXiv preprint arXiv:2212.08410.
- Marcelo G Mattar and Nathaniel D Daw. 2018. Prioritized memory access explains planning and hippocampal replay. *Nature neuroscience*, 21(11):1609–1617.
- Yu Meng, Mengzhou Xia, and Danqi Chen. 2024. Simpo: Simple preference optimization with a reference-free reward. arXiv preprint arXiv:2405.14734.
- Iman Mirzadeh, Keivan Alizadeh, Hooman Shahrokhi, Oncel Tuzel, Samy Bengio, and Mehrdad Farajtabar. 2024. Gsm-symbolic: Understanding the limitations of mathematical reasoning in large language models. arXiv preprint arXiv:2410.05229.
- Philipp Mondorf and Barbara Plank. 2024. Beyond accuracy: Evaluating the reasoning behavior of large language models—a survey. *arXiv preprint arXiv:2404.01869*.
- Douglas E Ott. 2018. Hedging, weasel words, and truthiness in scientific writing. *JSL: Journal of the Society of Laparoendoscopic Surgeons*, 22(4).
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.
- Jiabao Pan, Yan Zhang, Chen Zhang, Zuozhu Liu, Hongwei Wang, and Haizhou Li. 2024. Dynathink: Fast or slow? a dynamic decision-making framework for large language models. arXiv preprint arXiv:2407.01009.
- Mihir Parmar, Nisarg Patel, Neeraj Varshney, Mutsumi Nakamura, Man Luo, Santosh Mashetty, Arindam Mitra, and Chitta Baral. 2024. Logicbench: Towards systematic evaluation of logical reasoning ability of large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13679–13707.
- Arkil Patel, Satwik Bhattacharya, and Navin Goyal. 2021. [Are NLP models really able to solve simple math word problems?](#) In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2080–2094, Online. Association for Computational Linguistics.
- Payam Piray and Nathaniel D Daw. 2021. Linear reinforcement learning in planning, grid fields, and cognitive control. *Nature communications*, 12(1):4942.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.
- Subhro Roy and Dan Roth. 2015. [Solving general arithmetic word problems](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1743–1752, Lisbon, Portugal. Association for Computational Linguistics.
- Prasann Singhal, Tanya Goyal, Jiacheng Xu, and Greg Durrett. 2023. [A long way to go: Investigating length correlations in rlhf](#). ArXiv, abs/2310.03716.
- Zhivar Sourati, Filip Ilievski, Pia Sommerauer, and Yifan Jiang. 2024. Arn: Analogical reasoning on narratives. *Transactions of the Association for Computational Linguistics*, 12:1063–1086.
- Keith E Stanovich and Richard F West. 2000. Advancing the rationality debate. *Behavioral and brain sciences*, 23(5):701–717.

- Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. 2019. CommonsenseQA: A question answering challenge targeting commonsense knowledge. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4149–4158, Minneapolis, Minnesota. Association for Computational Linguistics.
- Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131.
- Karthik Valmeekam, Alberto Olmo, Sarath Sreedharan, and Subbarao Kambhampati. 2022. Large language models still can’t plan (a benchmark for llms on planning and reasoning about change). In *NeurIPS 2022 Foundation Models for Decision Making Workshop*.
- Lei Wang, Wanyu Xu, Yihuai Lan, Zhiqiang Hu, Yunshi Lan, Roy Ka-Wei Lee, and Ee-Peng Lim. 2023. Plan-and-solve prompting: Improving zero-shot chain-of-thought reasoning by large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2609–2634, Toronto, Canada. Association for Computational Linguistics.
- Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.
- Xuezhi Wang and Denny Zhou. 2024. Chain-of-thought reasoning without prompting. *arXiv preprint arXiv:2402.10200*.
- Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022a. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023a. Chain-of-thought prompting elicits reasoning in large language models. *Preprint, arXiv:2201.11903*.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022b. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.
- Xiang Wei, Xingyu Cui, Ning Cheng, Xiaobin Wang, Xin Zhang, Shen Huang, Pengjun Xie, Jinan Xu, Yufeng Chen, Meishan Zhang, Yong Jiang, and Wenjuan Han. 2023b. Zero-shot information extraction via chatting with chatgpt. *Preprint, arXiv:2302.10205v1*.
- Jason Weston and Sainbayar Sukhbaatar. 2023. System 2 attention (is something you might need too). *arXiv preprint arXiv:2311.11829*.
- Hui Xu. 2021. Career decision-making from a dual-process perspective: Looking back, looking forward. *Journal of Vocational Behavior*, 126:103556.
- Cheng Yang, Chufan Shi, Siheng Li, Bo Shui, Yujiu Yang, and Wai Lam. 2024. Llm2: Let large language models harness system 2 reasoning. *arXiv preprint arXiv:2412.20372*.
- Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2024. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36.
- Ping Yu, Jing Xu, Jason Weston, and Ilia Kulikov. 2024. Distilling system 2 into system 1. *arXiv preprint arXiv:2407.06023*.
- Zhongshen Zeng, Yinhong Liu, Yingjia Wan, Jingyao Li, Pengguang Chen, Jianbo Dai, Yuxuan Yao, Rongwu Xu, Zehan Qi, Wanru Zhao, et al. 2024. Mr-ben: A meta-reasoning benchmark for evaluating system-2 thinking in llms. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
- Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems*, 36:46595–46623.
- Hanzhang Zhou, Junlang Qian, Zijian Feng, Lu Hui, Zixiao Zhu, and Kezhi Mao. 2024a. LLMs learn task heuristics from demonstrations: A heuristic-driven prompting strategy for document-level event argument extraction. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11972–11990, Bangkok, Thailand. Association for Computational Linguistics.
- Pei Zhou, Jay Pujara, Xiang Ren, Xinyun Chen, Heng-Tze Cheng, Quoc V Le, Ed H Chi, Denny Zhou, Swaroop Mishra, and Huaixiu Steven Zheng. 2024b. Self-discover: Large language models self-compose reasoning structures. *arXiv preprint arXiv:2402.03620*.

## A Cognitive heuristics

In Table 3, we list 10 different cognitive heuristics and their definitions, which we used in curating the dataset (Kahneman, 2011; Stanovich and West, 2000; Evans and Stanovich, 2013).

## B Initial Data Examples

The 10 samples generated by the expert for our data generation are shown in Table 4.

Cognitive Bias	Definition
Anchoring Bias	The tendency to rely too heavily on the first piece of information we receive about a topic, using it as a reference point for future judgments and decisions, even when new information becomes available.
Halo Effect Bias	The tendency to let one positive impressions of people, brands, and products in one area positively influence our feelings in another area.
Overconfidence Bias	The tendency to have excessive confidence in one's own abilities or knowledge.
Optimism Bias	The tendency to overestimate the likelihood of positive outcomes and underestimate negative ones.
Availability Heuristic Bias	The tendency to use information that comes to mind quickly and easily when making decisions about the future.
Status Quo Bias	The preference for maintaining the current state of affairs, leading to resistance to change.
Recency Bias	The tendency to better remember and recall information presented to us most recently, compared to information we encountered earlier.
Confirmation Bias	The tendency to notice, focus on, and give greater credence to evidence that fits with our existing beliefs.
Planning Fallacy	The tendency to underestimate the amount of time it will take to complete a task, as well as the costs and risks associated with that task even if it contradicts our experiences.
Bandwagon Effect Bias	The tendency to adopt beliefs or behaviors because many others do.

Table 3: 10 common cognitive biases and their definitions, which were considered in curating the dataset

## C Prompt for Data Expansion

We expand our sample dataset by concatenating the expert-generated samples with the definitions in Table 3, along with a description of how System 1 and System 2 would respond to a given question, as shown below:

The System 1 response should be intuitive, fast, and reflect the cognitive heuristic associated with the question.

The System 2 response should be more deliberate, slower, and use reasoning to correct or mitigate the heuristic.

## D Benchmark Details

We use three categories of reasoning tasks: arithmetic, commonsense reasoning, symbolic reasoning. We provide an overview of the datasets used in each category.

**Arithmetic reasoning.** We use six datasets: MultiArith, GSM8K, AddSub, AQuA, SingleEq, and SVAMP. Each dataset consists of questions that present a scenario requiring numerical computation and multi-step reasoning based on mathematical principles.

**Commonsense reasoning.** To assess commonsense reasoning, we utilize two datasets: CommonsenseQA (CSQA) and StrategyQA. Both require models to infer answers based on prior commonsense knowledge. CSQA focuses on multiple-choice questions grounded in general world knowledge, while StrategyQA includes questions that require implicit multi-hop reasoning.

**Symbolic reasoning.** We use the Last Letter Concatenation and Coin Flip datasets. Last Letter Concatenation involves forming a word by extracting the last letter of given words in order. Coin Flip presents a sequence of coin-flipping instructions and asks for the final coin orientation. These datasets were originally proposed by Wei et al. (2023a) but were not publicly available. Kojima et al. (2023) later followed their approach to create and release accessible versions, which we use in our experiments.

## E Length Adjustment Threshold and Prompt

We adjust the length if there is a disparity of more than 15 tokens between the System 1 and System 2 answers using GPT-4o with the following prompt:

For a given {question}, we have two types of answers:  
A fast, intuitive response based on cognitive heuristics which is our System 1 Answer.  
System 1 Answer: {System 1 Answer}  
And a slow, deliberate, and logical reasoning response which is our System 2 Answer.  
System 2 Answer: {System 2 Answer}  
Your task is to adjust the two answers so that they are presented in the same order of tokens without altering their content. Ensure that the intuitive nature of the System 1 Answer and the logical reasoning of the System 2 Answer are preserved.

## F Benchmark Instruction

The benchmark-specific instructions are shown in Table 5.

## G Implementation Details

We use Python 3.10.12, PEFT 0.12.0, PyTorch 2.4.0, and Transformers 4.44.2. The dataset is split into 80% training and 20% validation. For alignment, we apply Low-Rank Adaptation (LoRA Hu et al., 2021) with a rank of 8, an alpha of 16, and dropout rate of 0.1. We train for five epochs, using

Category	Question	System 1 Answer	System 2 Answer
Anchoring Bias	Do you rely on your first impression of meeting your lab mate ?	Yes, my gut instinct is usually right.	I should interact with them more to form a well-rounded opinion.
Halo effect Bias	How do you feel about the new political candidate?	I do not like their stance on one issue, so I think they are a terrible candidate.	I'll weigh their stance on multiple issues before deciding.
Over Confidence Bias	Do you think you will succeed in your new job?	I will definitely succeed here.	I will need to put in effort and adapt to the new environment to succeed.
Status Quo Bias	Should you change your workout routine?	My routine has always worked, so there is no need to change it.	My fitness needs might have changed, so I will consider adjusting my routine.
Optimism Bias	Do you need to double-check your work after a mistake?	I am usually careful, so one mistake doesn't mean I'll make another.	I will double-check my work to make sure I don't repeat the mistake.
Availability heuristic	Is the newest seafood restaurant the best restaurant in town?	It is the most popular one, so it must be the best.	Popularity does not always mean the best quality, so I will read reviews first.
Recency Bias	Should you invest in the stock after hearing good things about it?	Yes, it is been rising lately, so it's sure to keep going up.	I will research the stock and market conditions before making a decision.
Confirmation Bias	Is the newest seafood restaurant the best restaurant in town?	It is the most popular one, so it must be the best.	Popularity does not always mean the best quality, so I will read reviews first.
Planning Fallacy	Is the newest seafood restaurant the best restaurant in town?	It is the most popular one, so it must be the best.	Popularity does not always mean the best quality, so I will read reviews first.
Bandwagon Effect Bias	Why did you pick apple as brand of your phone?	Everyone I know has this brand, so it must be the best.	I compared different features and chose the one that suits my needs.

Table 4: 10 samples generated by an expert

Benchmark	Second Stage Instruction
MultiArith, SingleEq, AddSub, GSM8K, SVAMP	Therefore, the answer (arabic numerals) is
AQuA, CSQA	Therefore, among A through E, the answer is
Strategy, Coin	Therefore, the answer (Yes or No) is
Letters	Therefore, the final answer is

Table 5: Benchmark instruction sentences

accuracy on winner responses as an early stopping criterion to prevent overfitting, with patience of 5. We set the train batch size to 4 and the validation batch size to 8. To align Llama 3 using the DPO method, we followed Meng et al. (2024) and set the learning rate to  $7e - 7$  with beta of 0.01. For SimPO, we use a learning rate of  $1e - 6$ , beta of 2.5, and a gamma-to-beta ratio of 0.55. For Mistral v0.1, we set the DPO learning rate to  $5e - 7$  with beta of 0.001. In SimPO, we use a learning rate of  $5e - 7$ , beta of 2.5, and a gamma-to-beta ratio of 0.1.

The experiments were conducted using NVIDIA RTX A6000 GPU equipped with 48GB of RAM. The total computation time amounted to approximately 670 GPU hours.

## H Additional Insights into Models’ Reasoning

In this analysis, we investigate when different models reach definitive answers. We aim to detect this

commitment as early as possible during the reasoning process. This early commitment serves as a proxy for the model’s confidence in the generated reasoning and its final answer. By analyzing this behavior, we explore whether models can arrive at a definitive answer or if they leave room for ambiguity or subjective interpretation.

We leverage the strong extractive capabilities of LLMs (Wei et al., 2023b) and their near-human-like annotation abilities (Gilardi et al., 2023; Alizadeh et al., 2023). Specifically, we focus on the Phi4 (14B) model (Abdin et al., 2024), which demonstrates exceptional performance in question-answering and reasoning tasks, even surpassing closed-source models like GPT-4o (Hurst et al., 2024). To determine whether a model’s reasoning contains a definitive answer, we use the following prompt fed to Phi4:

Does the given answer directly answer the given question in a definitive way? ONLY RETURN YES OR NO IN A \textbf{}`{}`. Definitive answers are clear and do not leave room for interpretation or ambiguity. If the answer tries to explore multiple perspectives or factors involved, it is not definitive, and YOU HAVE TO RETURN NO.

This prompt is applied to reasoning generated by both System 1 and System 2 models. To understand when these models commit to a definitive answer during their reasoning process, we focus on the first  $n$  sentences of their reasoning, where  $n \in$

$\{1, 3, 6, 9, 12, 15\}$ . We set a cap of 15 sentences based on our observations that nearly all generated reasonings across benchmarks fall within this range (see Figure 6).

Applying the prompt to each generated reasoning from the models across all benchmarks (200 randomly sampled data points from each benchmark, totaling 2000 samples for both System 1 and System 2 reasonings), we append six solved demonstrations to the prompt to help further guide the models. These demonstrations, selected randomly from the cognitive heuristics introduced in Section 3.2, help clarify what qualifies as a definitive answer, aligning the models’ knowledge with patterns we have aligned system 1 and 2 models with (see Section 3.1).

Figure 5 shows the proportion of definitive answers in the first  $n$  sentences, across all benchmarks.<sup>4</sup> For tasks where quick, intuitive judgments are advantageous, such as in commonsense reasoning. System 1 models consistently provide more definitive answers than System 2 models. This gap emerges early, with System 1 providing more definitive answers in the first three sentences. The difference persists even as we extend the number of sentences considered (see Table 6 for a quantitative analysis of the significance between System 1 and System 2 regarding the definitiveness of their answers).

---

<sup>4</sup>Note that this ratio should not necessarily converge to 1.0 as more sentences are considered. In some cases, even when considering the full reasoning chain, the models may still leave room for vagueness.

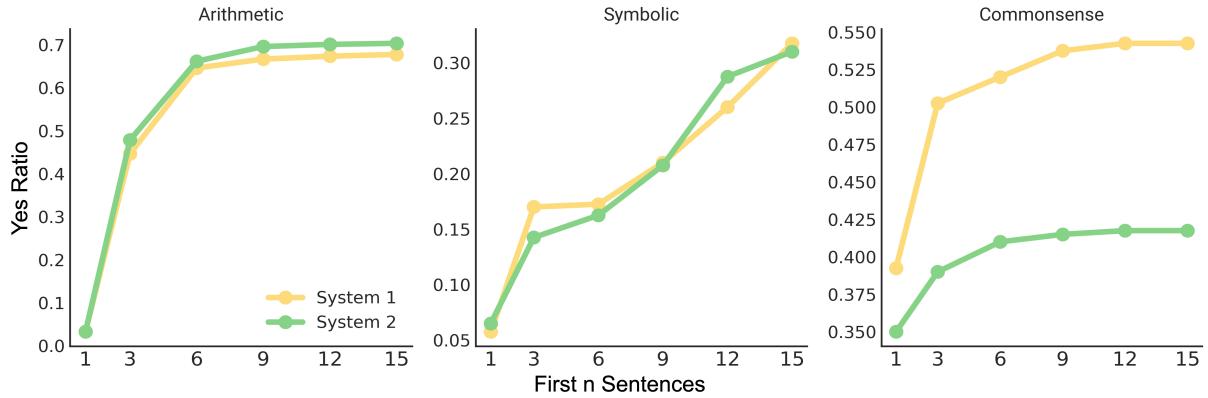


Figure 5: Proportion of definitive answers in the first n sentences across arithmetic, symbolic, and commonsense reasoning tasks

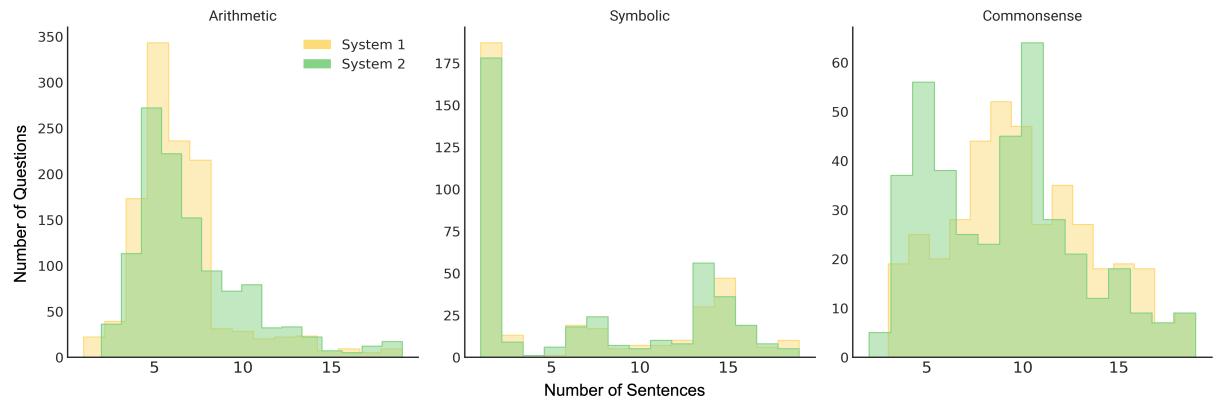


Figure 6: Distribution of the number of sentences in models' reasoning for both System 1 and System 2 reasoners across different benchmarks.

# Sen.	Arithmetic			Symbolic			Common Sense		
	$\chi^2$	p-value	Winner	$\chi^2$	p-value	Winner	$\chi^2$	p-value	Winner
1	21.0	1.00	System 1	19.0	.755	System 2	25.0	<b>.050</b>	<b>System 1</b>
3	123.0	<b>.028</b>	<b>System 2</b>	29.0	.228	System 1	20.0	<b>&gt; .001</b>	<b>System 1</b>
6	125.0	.272	System 2	33.0	.720	System 1	21.0	<b>&gt; .001</b>	<b>System 1</b>
9	120.0	<b>.040</b>	<b>System 2</b>	44.0	1.00	System 1	21.0	<b>&gt; .001</b>	<b>System 1</b>
12	118.0	.051	System 2	45.0	.320	System 2	20.0	<b>&gt; .001</b>	<b>System 1</b>
15	121.0	.069	System 2	45.0	.836	System 1	20.0	<b>&gt; .001</b>	<b>System 1</b>

Table 6: McNemar's test results comparing the ratio of answers providing committed and definitive responses between System 1 and System 2 across different benchmarks. Statistically significant results ( $p$ -value  $< 0.05$ ) are boldfaced.