Hallucination Mitigation in Reasoning Models

Motivation

Following the emergence of DeepSeek and GPT-5, reasoning-oriented large language models have attracted significant research attention. DeepSeek's remarkable performance stems from skipping the conventional supervised fine-tuning stage and instead employing pure GRPO-based reinforcement learning, which fundamentally enhances reasoning optimization through direct policy refinement. However, this approach has also been observed to increase the frequency of hallucinations—broadly referring to the model's tendency to generate incorrect or unsupported responses. This project aims to investigate potential mitigation strategies by refining chain-of-thought fine-tuning and reward-model design to reduce hallucination occurrence while preserving reasoning quality.

Planned Work

Overview

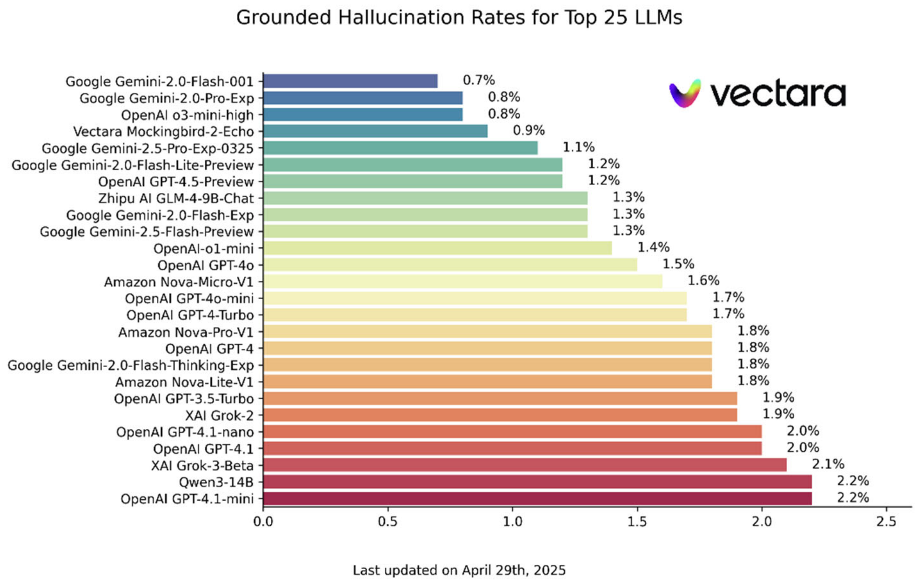
This project aims not only to evaluate but also to **mitigate hallucinations in reasoning-oriented LLMs** through targeted optimization of the **reward model** and **CoT fine-tuning**. Specifically, we will investigate how adjusting **PRM** and **ORM** within the **GRPO training framework** affects hallucination behavior. By dynamically balancing these components and identifying hallucination onset stages during multi-step reasoning, we aim to develop a mechanism that both detects and suppresses model "fabrication" tendencies while maintaining reasoning depth.

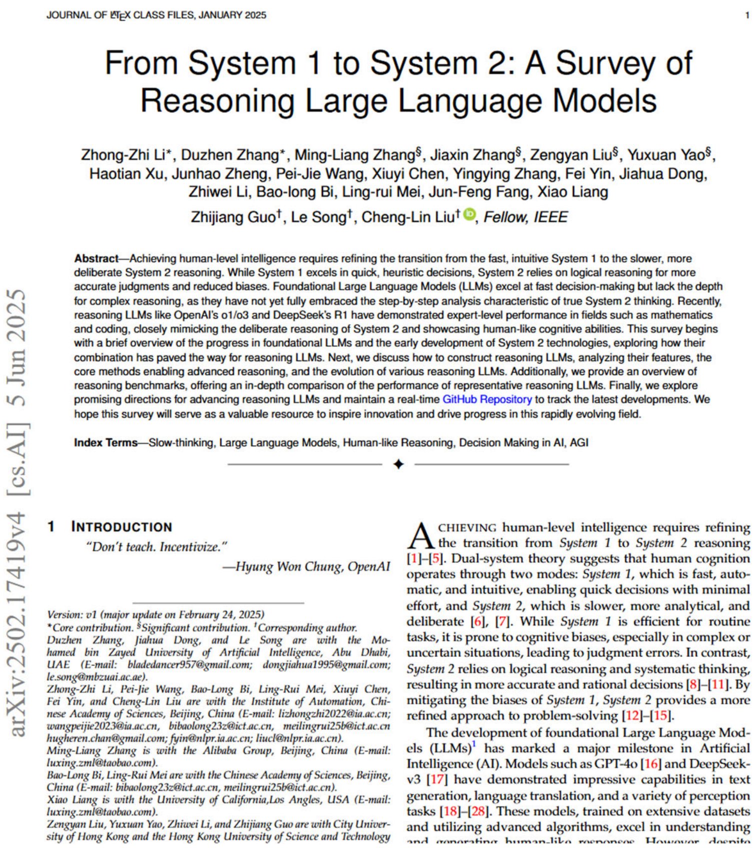
Methodology and Experimental Plan

The study will proceed in structured phases, combining model evaluation, dataset expansion, and iterative ablation experiments:

Phase	Description
0. Define hallucination & test scope	Establish a precise operational definition of hallucination, classify subtypes (perceptual, delusional, confabulatory, etc.), and delimit the evaluation scope across reasoning benchmarks.
1. Small-scale model + benchmark evaluation	Conduct pilot runs using a compact model to validate the dataset, prompts, and automated scoring workflow.
2. Data expansion & model improvement	Expand and refine the dataset, perform fine-tuning, and quantify improvements in hallucination reduction.
3. Cross-benchmark validation	Test the refined approach on open-source datasets using distributed evaluation to ensure generality.
4. Reward-model adjustment & ablation	Adjust PRM and ORM weightings within GRPO; analyze their effects on factual accuracy and reasoning coherence.
5. Further ablation & exploratory analysis	Perform statistical testing, case analysis, and visualize hallucination emergence during chain-of-thought reasoning.

Reference:





JOURNAL OF LATEX CLASS FILES, JANUARY 2025 1

From System 1 to System 2: A Survey of Reasoning Large Language Models

Zhong-Zhi Li*, Duzhen Zhang*, Ming-Liang Zhang§, Jiaxin Zhang§, Zengyan Liu§, Yuxuan Yao§, Haotian Xu, Junhao Zheng, Pei-Jie Wang, Xiuyi Chen, Yingying Zhang, Fei Yin, Jiahua Dong, Zhiwei Li, Bao-long Bi, Ling-rui Mei, Jun-Feng Fang, Xiao Liang

Zhijiang Guo†, Le Song†, Cheng-Lin Liu†, *Fellow, IEEE*

Abstract—Achieving human-level intelligence requires refining the transition from the fast, intuitive System 1 to the slower, more deliberate System 2 reasoning. While System 1 excels in quick, heuristic decisions, System 2 relies on logical reasoning for more accurate judgments and reduced biases. Foundational Large Language Models (LLMs) excel at fast decision-making but lack the depth for complex reasoning, as they have not yet fully embraced the step-by-step analysis characteristic of true System 2 thinking. Recently, reasoning LLMs like OpenAI's o1/o3 and DeepSeek's R1 have demonstrated expert-level performance in fields such as mathematics and coding, closely mimicking the deliberate reasoning of System 2 and showcasing human-like cognitive abilities. This survey begins with a brief overview of the progress in foundational LLMs and the early development of System 2 technologies, exploring how their combination has paved the way for reasoning LLMs. Next, we discuss how to construct reasoning LLMs, analyzing their features, the core methods enabling advanced reasoning, and the evolution of various reasoning LLMs. Additionally, we provide an overview of reasoning benchmarks, offering an in-depth comparison of the performance of representative reasoning LLMs. Finally, we explore promising directions for advancing reasoning LLMs and maintain a real-time GitHub Repository to track the latest developments. We hope this survey will serve as a valuable resource to inspire innovation and drive progress in this rapidly evolving field.

Index Terms—Slow-thinking, Large Language Models, Human-like Reasoning, Decision Making in AI, AGI

arXiv:2502.17419v4 [cs.AI] 5 Jun 2025

1 Introduction

"Don't teach. Incentivize."

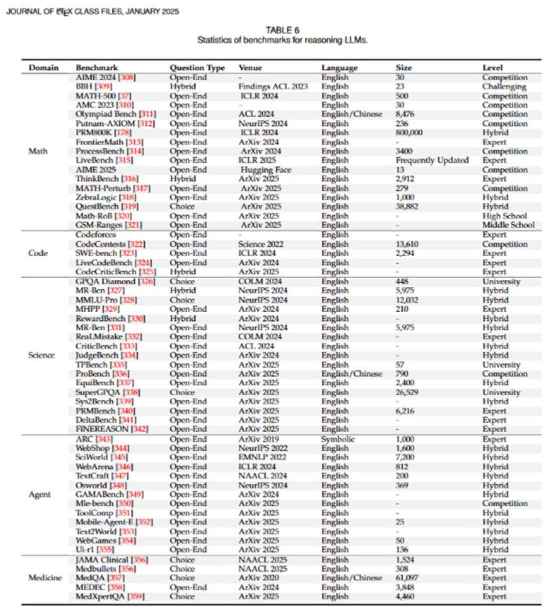
—Hyung Won Chung, OpenAI

Achieving human-level intelligence requires refining the transition from *System 1* to *System 2* reasoning [1]–[5]. Dual-system theory suggests that human cognition operates through two modes: *System 1*, which is fast, automatic, and intuitive, enabling quick decisions with minimal effort, and *System 2*, which is slower, more analytical, and deliberate [6], [7]. While *System 1* is efficient for routine tasks, it is prone to cognitive biases, especially in complex or uncertain situations, leading to judgment errors. In contrast, *System 2* relies on logical reasoning and systematic thinking, resulting in more accurate and rational decisions [8]–[11]. By mitigating the biases of *System 1*, *System 2* provides a more refined approach to problem-solving [12]–[15].

The development of foundational Large Language Models (LLMs)[1] has marked a major milestone in Artificial Intelligence (AI). Models such as GPT-4o [16] and DeepSeek-v3 [17] have demonstrated impressive capabilities in text generation, language translation, and a variety of perception tasks [18]–[28]. These models, trained on extensive datasets and utilizing advanced algorithms, excel in understanding

Version: v1 (major update on February 24, 2025)
**Core contribution. §Significant contribution. †Corresponding author.*
Duzhen Zhang, Jiahua Dong, and Le Song are with the Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE (E-mail: bladedancer957@gmail.com; dongjiahua1995@gmail.com; le.song@mbzuai.ac.ae).
Zhong-Zhi Li, Pei-Jie Wang, Bao-Long Bi, Ling-Rui Mei, Xiuyi Chen, Fei Yin, and Cheng-Lin Liu are with the Institute of Automation, Chinese Academy of Sciences, Beijing, China (E-mail: lizhongzhi2022@ia.ac.cn; wangpeijie2023@ia.ac.cn, bibaolong23z@ict.ac.cn, meilingrui25b@ict.ac.cn hugheren.chan@gmail.com; fyin@nlpr.ia.ac.cn; liucl@nlpr.ia.ac.cn).
Ming-Liang Zhang is with the Alibaba Group, Beijing, China (E-mail: luxing.zml@taobao.com).
Bao-Long Bi, Ling-Rui Mei are with the Chinese Academy of Sciences, Beijing, China (E-mail: bibaolong23z@ict.ac.cn, meilingrui25b@ict.ac.cn).
Xiao Liang is with the University of California,Los Angles, USA (E-mail: luxing.zml@taobao.com).
Zengyan Liu, Yuxuan Yao, Zhiwei Li, and Zhijiang Guo are with City University of Hong Kong and the Hong Kong University of Science and Technology

https://arxiv.org/abs/2502.17419



JOURNAL OF LATEX CLASS FILES, JANUARY 2025 14

TABLE 6
Statistics of benchmarks for reasoning LLMs.

Domain	Benchmark	Question Type	Venue	Language	Size	Level
Math	AIME 2024 [308]	Open-End	-	English	30	Competition
	BBH [309]	Hybrid	Findings ACL 2023	English	23	Challenging
	MATH-500 [37]	Open-End	ICLR 2024	English	500	Competition
	AMC 2023 [310]	Open-End	–	English	30	Competition
	Olympiad Bench [311]	Open-End	ACL 2024	English/Chinese	8,476	Competition
	Putnam-AXIOM [312]	Open-End	NeurIPS 2024	English	236	Competition
	PRM800K [178]	Open-End	ICLR 2024	English	800,000	Hybrid
	FrontierMath [313]	Open-End	ArXiv 2024	English	-	Expert
	ProcessBench [314]	Open-End	ArXiv 2024	English	3400	Competition
	LiveBench [315]	Open-End	ICLR 2025	English	Frequently Updated	Expert
	AIME 2025	Open-End	Hugging Face	English	13	Competition
	ThinkBench [316]	Hybrid	ArXiv 2025	English	2,912	Expert
	MATH-Perturb [317]	Open-End	ArXiv 2025	English	279	Competition
	ZebraLogic [318]	Open-End	ArXiv 2025	English	1,000	Hybrid
	QuestBench [319]	Choice	ArXiv 2025	English	38,882	Hybrid
	Math-RoB [320]	Open-End	ArXiv 2025	English	-	High School
	GSM-Ranges [321]	Open-End	ArXiv 2025	English	-	Middle School
Code	Codeforces	Open-End	-	English	-	Expert
	CodeContests [322]	Open-End	Science 2022	English	13,610	Competition
	SWE-bench [323]	Open-End	ICLR 2024	English	2,294	Expert
	LiveCodeBench [324]	Open-End	ArXiv 2024	English	-	Expert
	CodeCriticBench [325]	Hybrid	ArXiv 2025	English	-	Expert
Science	GPQA Diamond [326]	Choice	COLM 2024	English	448	University
	MR-Ben [327]	Hybrid	NeurIPS 2024	English	5,975	Hybrid
	MMLU-Pro [328]	Choice	NeurIPS 2024	English	12,032	Hybrid
	MHPP [329]	Open-End	ArXiv 2024	English	210	Expert
	RewardBench [330]	Hybrid	ArXiv 2024	English	-	Hybrid
	MR-Ben [331]	Open-End	NeurIPS 2024	English	5,975	Hybrid
	RealMistake [332]	Open-End	COLM 2024	English	-	Expert
	CriticBench [333]	Open-End	ACL 2024	English	-	Hybrid
	JudgeBench [334]	Open-End	ArXiv 2024	English	-	Hybrid
	TPBench [335]	Open-End	ArXiv 2025	English	57	University
	ProBench [336]	Open-End	ArXiv 2025	English/Chinese	790	Competition
	EquiBench [337]	Open-End	ArXiv 2025	English	2,400	Hybrid
	SuperGPQA [338]	Choice	ArXiv 2025	English	26,529	University
	Sys2Bench [339]	Open-End	ArXiv 2025	English	-	Hybrid
	PRMBench [340]	Open-End	ArXiv 2025	English	6,216	Expert
	DeltaBench [341]	Open-End	ArXiv 2025	English	-	Expert
	FINEREASON [342]	Open-End	ArXiv 2025	English	-	Expert
Agent	ARC [343]	Open-End	ArXiv 2019	Symbolic	1,000	Expert
	WebShop [344]	Open-End	NeurIPS 2022	English	1,600	Hybrid
	SciWorld [345]	Open-End	EMNLP 2022	English	7,200	Hybrid
	WebArena [346]	Open-End	ICLR 2024	English	812	Hybrid
	TextCraft [347]	Open-End	NAACL 2024	English	200	Hybrid
	Osworld [348]	Open-End	NeurIPS 2024	English	369	Hybrid
	GAMABench [349]	Open-End	ArXiv 2024	English	-	Hybrid
	Mle-bench [350]	Open-End	ArXiv 2025	English	-	Competition
	ToolComp [351]	Open-End	ArXiv 2025	English	-	Hybrid
	Mobile-Agent-E [352]	Open-End	ArXiv 2025	English	25	Hybrid
	Text2World [353]	Open-End	ArXiv 2025	English	-	Hybrid
	WebGames [354]	Open-End	ArXiv 2025	English	50	Hybrid
	Ui-r1 [355]	Open-End	ArXiv 2025	English	136	Hybrid
Medicine	JAMA Clinical [356]	Choice	NAACL 2025	English	1,524	Expert
	Medbullets [356]	Choice	NAACL 2025	English	308	Expert
	MedQA [357]	Choice	ArXiv 2020	English/Chinese	61,097	Expert
	MEDEC [358]	Open-End	ArXiv 2024	English	3,848	Expert
	MedXpertQA [359]	Choice	ArXiv 2025	English	4,460	Expert