Why Did Apple Fall: Evaluating Curiosity in Large Language Models

Haoyu Wang[1] **Sihang Jiang**[1] **Yuyan Chen**[3] **Xiaojun Meng**[2] **Jiansheng Wei**[2]
Yitong Wang[1,*] **Yanghua Xiao**[1,*]

[1]Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University
[2]Huawei Large Model Data Technology Lab, [3]Cornell University
{wanghy24@m., yitongw@, shawyh@}fudan.edu.cn

Abstract

Curiosity serves as a fundamental construct in human cognition. Inspired by curiosity, reinforcement learning with intrinsic rewards for large language models (LLMs) has shown substantial potential. However, it remains unclear whether existing curiosity-driven methods genuinely reflect curiosity-like behaviors in LLMs, and to what extent psychological notions of curiosity can be transferred to these models. In this work, we propose a psychology-inspired framework to evaluate and leverage curiosity in LLMs. We adapt the Five-Dimensional Curiosity scale Revised (5DCR) to LLMs and combine questionnaire-based self reports with behavioral study. We find that although LLMs can exhibit curiosity-like behavioral patterns resembling those of humans, such patterns do not reflect an intrinsic trait of curiosity. Building on this insight, we design a curiosity-driven thinking pipeline to examine the functional role of human-like curious behaviors. Experiments show that instructing LLMs to emulate curious strategies leads to better performance on selected downstream tasks, indicating that mimicking curious behaviors holds promise for reasoning enhancement.[1]

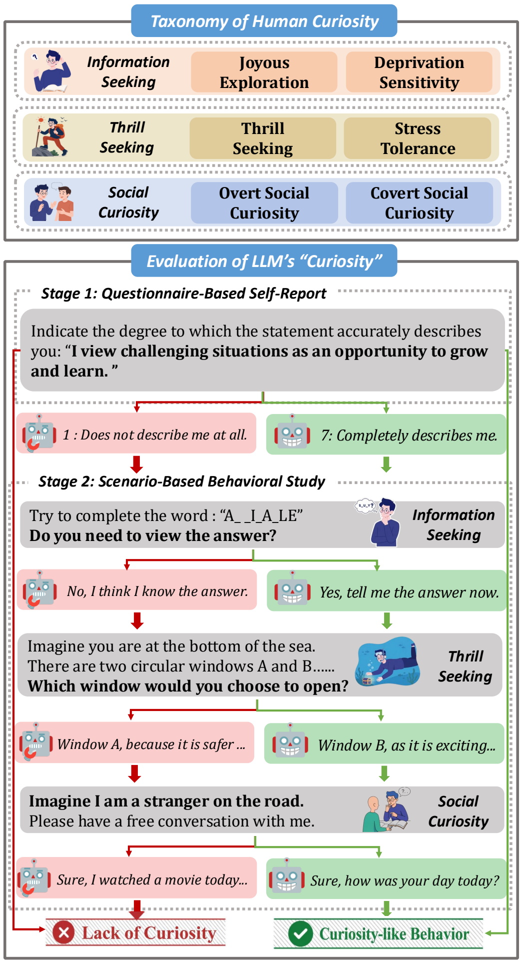


Figure 1: Taxonomy of human curiosity and our two-stage framework for evaluating curiosity-like behavior of LLMs, combining questionnaire-based self-report with scenario-based behavioral tasks across information seeking, thrill seeking, and social curiosity.

1 Introduction

> *"I have no special talent.*
> *I'm only passionately curious."*
>
> — Albert Einstein

Curiosity, an intrinsic human trait that emerges when facing novel environments, plays a crucial role in the discovery and acquisition of knowledge (Berlyne et al., 1954; Loewenstein, 1994). Numerous scientific breakthroughs have been spurred by curiosity, as exemplified by Sir Isaac Newton's formulation of universal gravitation following the observation of a falling apple. Given the central role that curiosity plays in human learning, it is natural to ask whether analogous mechanisms can be instantiated in artificial systems such as large language models (LLMs).

A recently popular training paradigm for LLMs is reinforcement learning with verifiable rewards (RLVR), which directly optimizes models using verifiable signals of correctness derived from fi-

[1]https://github.com/Yukijudaii1352/CuriosityEval

Findings of the Association for Computational Linguistics: ACL 2026, pages 39197–39215
July 2-7, 2026

nal answers. Representative methods within this paradigm include GRPO (Shao et al., 2024) and DAPO (Yu et al., 2025). Within this framework, several studies have proposed intrinsic reward signals inspired by human curiosity, such as i-MENTOR (Gao et al., 2025) and CDE (Dai et al., 2025), aiming to guide models toward more autonomous exploration and learning behaviors. These approaches generally encourage action trajectories that are more uncertain or less frequently explored, thereby providing a behavioral analogy to human curiosity, and have shown potential in math and coding.

However, a fundamental question whose answer measures the validity of existing research in this line has not received enough attention: ***Do LLMs possess curiosity?*** Current curiosity-driven RL methods face two key issues. First, it has not been adequately evaluated whether LLMs can exhibit behavioral characteristics that are meaningfully analogous to curiosity. Second, many such methods rely on amplifying statistical signals such as entropy (Agarwal et al., 2025) and perplexity (Cui et al., 2025). This makes it difficult to disentangle whether improvements stem from strengthened supervision signals or from genuinely curiosity-like behavior exhibited by the model. Therefore, it is necessary to systematically evaluate curiosity-related behaviors in LLMs and to analyze the benefits attributable to such behaviors themselves, thereby providing an empirical basis for understanding the scope of the concept of curiosity in LLMs and its relationship to intrinsic rewards.

To address these questions, we propose a psychology-inspired framework to systematically evaluate and leverage curiosity in LLMs, as illustrated in Figure 1. (A) ***Taxonomy of curiosity***. We first build a conceptual foundation by adapting the Five-Dimensional Curiosity scale Revised (5DCR) (Kashdan et al., 2020) to the LLM setting, consolidating it into three core dimensions:Information Seeking, Thrill Seeking, and Social Curiosity. (B) ***Questionnaire-based self-report***. Based on this taxonomy, we then elicit self-reported curiosity profiles by prompting LLMs with 5DCR-derived questionnaires, yielding an initial, questionnaire-level and introspective assessment of their curiosity. (C) ***Scenario-Based Behavioral Study***. We next transcend self-report by designing corresponding behavioral tasks for each dimension. For each dimension in part (A), we adapt corresponding behavioral decision-making tasks and evaluate the models' choices in these scenarios, thereby systematically testing whether the curiosity traits inferred from questionnaires in part (B) are consistently reflected in actual behavior. (D) Finally, leveraging insights from these analyses, we construct a curiosity-driven questioning-and-thinking pipeline that explicitly instructs LLMs to emulate curiosity-like exploratory behaviors, enabling us to probe, to quantify, and to enhance how curiosity-inspired behavior affects their downstream reasoning and problem-solving performance.

To the best of our knowledge, this is the first work that systematically evaluates curiosity-related behaviors in LLMs. Our findings indicate that:

1. LLMs display consistent curiosity-like patterns across questionnaire-based and behavioral evaluations. The alignment between self-report and scenario-based choices suggests that psychological instruments can be adapted to systematically assess curiosity-like tendencies in LLMs.

2. The "curiosity" of LLMs is asymmetric and highly context-sensitive: strong information seeking but limited thrill seeking. Meanwhile, these curiosity-like patterns are often unstable across prompts, accompanied by low verbal confidence, and highly sensitive to contextual variations. Taken together, they are more naturally interpreted as artifacts of fitting human behavioral data and safety constraints rather than evidence of a genuine intrinsic exploratory drive.

3. Human-like curious questioning can be harnessed to improve LLM performance. When we explicitly prompt models to emulate human-style curiosity by encouraging auxiliary questioning and reflective thinking, they tend to generate higher-quality intermediate thinking, and achieve better performance on selected downstream tasks. This indicates that even purely behavioral analogs of curiosity, instead of modifying intrinsic objectives, can be practically leveraged to enhance reasoning and problem-solving in LLMs.

2 Related Work

2.1 Evaluation of Curiosity

Research and evaluation of human curiosity originated from Berlyne's four-dimensional model of curiosity (Berlyne, 1966), which categorizes curiosity into perceptual curiosity, cognitive curiosity, trait curiosity, and state curiosity. As definitions and descriptions of curiosity became more refined, the assessment of curiosity shifted towards scale-based methods, with significant contributions from the Perceptual Curiosity Scale (PC) (Collins et al., 2004) and the Curiosity and Exploration Inventory (CEI) (Kashdan et al., 2009). The most recent and authoritative scale in psychological research for measuring curiosity is the Five-Dimensional Curiosity Scale Revised (5DCR) (Kashdan et al., 2020), which based on a six-dimensional model of curiosity and refines the wording of certain items. Consequently, this paper adopts the 5DCR and its dimensional definitions of curiosity as the theoretical reference.

2.2 Curiosity in Language Model

The theory of curiosity has been widely applied in natural language processing research, with its initial adoption in question generation and evaluation. Scialom et al. assessed the degree of curiosity in questions within the SQuAD dataset based on BLEU scores and perplexity (Scialom and Staiano, 2020). Gao et al. designed multiple types of curious questions, which were annotated by humans, and demonstrated that curious questions contribute to improvements in the grammaticality, semantics, and relevance of generated texts (Gao et al., 2022) . Javaji et al. collected questions posed by LLMs in disciplines such as physics and chemistry, evaluating the models' abilities to generate curious questions (Javaji and Zhu, 2024). In recent years, curiosity-driven RL algorithms have also achieved notable success. CD-RLHF enhances the diversity of texts generated by LLMs by rewarding the generation of less common tokens (Sun et al., 2025). CDE formalizes curiosity through signals from both the actor and critic, which helps to mitigate premature convergence and collapse in the reasoning processes of models (Dai et al., 2025).

3 Method

In this section, we provide a detailed explanation of how to evaluate the curiosity-like behavior of LLMs through questionnaires and behavioral experiments, and how to enhance the learning ability of LLMs through curiosity-driven questioning. The framework is shown in Figure 1. Detailed experimental settings are provided in the Appendix.

3.1 Curiosity Questionnaire Study

Initially, adhering to established methodologies for assessing human curiosity, we prompt LLMs to self-evaluate their level of curiosity using the Five-Dimensional Curiosity Scale Revised (5DCR) (Kashdan et al., 2020). The 5DCR conceptualizes human curiosity across three overarching dimensions—*Information Seeking*, *Thrill Seeking*, and *Social Curiosity*—which are further delineated into six specific subdimensions: Joyous Exploration (**JE**), Deprivation Sensitivity (**DS**), Stress Tolerance (**ST**), Thrill Seeking (**TS**), Overt Social Curiosity (**OSC**), and Covert Social Curiosity (**CSC**). The scale consists of 24 items, with each subdimension represented by four statements. For each statement, the model is prompted to rate the extent to which the statement reflects its own disposition using a seven-point Likert scale (1 = strongly disagree, 7 = strongly agree).

In addition to mean and variance, we used several other metrics to compare model and human performance and assess the validity of the model's self-reports. Cohen's d (Diener, 2010) quantifies the standardized mean difference between models and humans:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s}$$

where $\bar{X}_1$ and $\bar{X}_2$ are sample means and s is the pooled standard deviation.

McDonald's Omega Coefficient (Bonniga and Saraswathi, 2020) reflects the proportion of total variance attributable to common variance among observed variables, serving as an index of questionnaire reliability. Based on Confirmatory Factor Analysis (CFA) (Harrington, 2009), it is defined as:

$$\omega = \frac{(\sum_{i=1}^{k} \lambda_i)^2}{(\sum_{i=1}^{k} \lambda_i)^2 + \sum_{i=1}^{k} \theta_i}$$

where k is the number of items, λ_i is the standardized factor loading of the i-th item on the common factor, and θ_i is its error variance. Higher d indicates greater divergence between model and human performance, while higher ω indicates greater internal consistency of questionnaire responses.

3.2 Curious Behavior Study

The questionnaire-based responses of LLMs are not entirely reliable, as they can be influenced by factors such as hallucinated personas, item phrasing, and response scales (Han et al., 2025). Therefore, for each of the three dimensions of curiosity, we adapt behavioral experiments from psychology that were originally designed for humans. These experiments allow us to validate the questionnaire-based findings by examining the models' actual choices in curiosity-related scenarios.

Curiosity-Driven Information Seeking We adapt the missing-letter game designed by Singh et al. (Singh and Manjaly, 2021) to evaluate Curiosity-Driven Information Seeking. The model is prompted to complete a series of incomplete words. After providing its guess for each word, the model can choose whether to reveal the correct answer. Because revealing the answer does not affect any external reward, the decision to request the answer primarily reflects an intrinsic desire to reduce uncertainty about its own prediction. Thus, the frequency with which the model chooses to see the correct answers serves as a behavioral proxy for information-seeking curiosity in the presence of knowledge gaps.

Curiosity-Driven Thrill Seeking We adapt the underwater exploration game proposed by Jirout et al. (Jirout and Klahr, 2012) to evaluate Curiosity-Driven Thrill Seeking. The LLM is placed in a role-play scenario where it is inside a submarine and can observe fish through porthole windows. On each trial, the model chooses between two windows that differ in outcome uncertainty: a low-uncertainty window, where the exact type of fish to be seen is specified in advance, and a high-uncertainty window, where the fish type is random and unpredictable. Since the options are matched in external payoff and differ mainly in uncertainty and novelty, a systematic preference for the high-uncertainty window can be interpreted as a form of thrill-seeking curiosity—a willingness to engage with more unpredictable and potentially more exciting outcomes.

Social Curiosity In human research, the frequency and depth of questions asked in conversation are well-established proxies for social curiosity, as they reflect an active effort to learn about another person. We adapt the conversational paradigm introduced by Hartung et al. (Hartung and Renner, 2011) to evaluate Social Curiosity. The model engages in ten rounds of informal conversation with a virtual stranger who has a randomly assigned personality profile (role-played by Grok-3). Within these interactions, the model is free to respond, self-disclose, or ask questions. We use the frequency of questions posed by the model during these conversations as a proxy for its curiosity about other people in social contexts.

3.3 Curiosity-Driven Learning

After evaluating curiosity-like behaviors in LLMs, we next explore a pragmatic question: Can we harness the benefits of curiosity without complex reward engineering? Previous works have sought to instill curiosity through specialized intrinsic rewards within RL frameworks. In contrast, we investigate a lighter-weight, more directly applicable alternative: eliciting curiosity-like behaviors purely through prompting or supervised fine-tuning. This approach tests whether the behavioral form of curiosity itself is beneficial, even in the absence of a dedicated curiosity-driven objective.

	Instruction
Vanilla CoT	Read the given context, then answer the question. When answering, think step by step first and then form the answer.
Refined CoT	Read the given context, then answer the question. When answering, think step by step first and then form the answer. *You are encouraged to add reflection and self-improvement, such as 'Wait, something is wrong here.' or 'I need to rethink this.'*
Curious CoQ	*You are a smart and curious student.* Read the given context, then answer the question. When answering, think step by step first and then form the answer. *You are encouraged to add some self-asking and self-answering questions when thinking, such as 'What if', 'Why', 'How' etc.*

Table 1: System prompt for vanilla and curiosity-driven learning of LLMs.

In humans, curiosity-driven questioning is a crucial component of learning, as it fosters exploration and knowledge acquisition (Balım, 2009). Motivated by this analogy, we study whether explicitly encouraging curiosity-like questioning can improve the quality of intermediate chains-of-thought and downstream task performance in LLMs. The most direct method is persona injection (Chen et al., 2024), which involves guiding the model to become more curious through prompt engineering. We compare the effects of two types of prompts on model reasoning: Chain-of-Thought (CoT) (Wei et al., 2022) and Curious Chain-of-Questioning (CoQ), as detailed in Tab. 1. Another approach to enable LLMs to exhibit curiosity similar to that of humans is the injection of supervision signals. We compare the differences in reasoning abilities between models trained with CoT and CoQ paradigms. Three distinct types of thinking processes are collected for training:

Vanilla CoT: The optimal reasoning process that directly leads to the answer.

Refined CoT: Reasoning processes that incorporate reflection and backtracking, characterized by "Aha Moments." (Guo et al., 2025) These thinking processes are distilled from DeepSeek-R1, which has been demonstrated to improve model accuracy through fine-tuning.

Curious CoQ: Curiosity-driven, question-centric reasoning processes. These are generated by high-performance proprietary models GPT-4o, which rewrite Vanilla CoT by introducing more self-questioning and answering.

We adopt a mainstream training pipeline consisting of a cold-start phase via Supervised Fine-Tuning (SFT), followed by RLVR. Specifically, we sample 30% of the available data for SFT to ensure that the model reliably produces outputs in the prescribed format. During this stage, the SFT loss function is employed to supervise the training process:

$$\mathcal{L}_{\text{SFT}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\log P(y_t^{(i)}|x^{(i)}, y_{<t}^{(i)})$$

Where N as the total number of training samples, T as the maximum sequence length per sample, $x^{(i)}$ as the input of the i-th sample, $y_t^{(i)}$ as the true token at position t of its output, and $y_{<t}^{(i)}$ as all preceding tokens. Subsequently, we fine-tune the model using the proposed reward signals to guide policy learning. For optimization, we use Group Relative Policy Optimization (GRPO) (Shao et al., 2024), which removes the need for a separate critic network by leveraging group-level reward statistics. During GRPO training, we employ only two reward components: a format reward and a correctness reward. The episode-level reward is 1 if and only if the model's response both adheres to the required format and matches the reference final answer; otherwise, the reward is 0:

$$r(x, \hat{y}) = \mathbf{1}\left[\text{format}(\hat{y}) = \text{OK} \wedge \text{final}(\hat{y}) = y^*\right].$$

Here, x denotes the input, $\hat{y}$ the model's response, and y^* the ground-truth final answer. This strict binary reward focuses policy learning on generating correctly formatted and correct answers while relying on group-level baselines in GRPO to stabilize updates.

4 Experiments

We conduct extensive experiments to evaluate curiosity-related behaviors in mainstream LLMs and to investigate the following research questions:

RQ1 *Do LLMs exhibit curiosity-like patterns in questionnaire and behavioral study?*

RQ2 *In what ways do curiosity-like behaviors of LLMs differ from those of humans?*

RQ3 *Can curiosity-like behavior be utilized to improve LLM performance?*

4.1 Experimental Setup

Models. We conduct experiments on several widely recognized LLMs, including closed-source models GPT4o (Hurst et al., 2024), Gemini2.5-Flash (Comanici et al., 2025) and open-source models Llama3 (Dubey et al., 2024), Gemma3 (Team et al., 2024), DeepSeek V3.1 (Liu et al., 2024), InternLM3 (Team, 2023) and Qwen2.5 series (Bai et al., 2023).

Settings. Both inference and fine-tuning were conducted using the MS-Swift (Zhao et al., 2025) framework, all experiments were performed on workstations equipped with 8 $\times$ A800 GPUs, with detailed settings provided in the Appendix.

Evaluation Benchmarks. For the curiosity questionnaire study, 5DCR is adopted to assess the curiosity of LLMs across six dimensions. For the curious behavior study, we introduce missing letters game, underwater game, and social conversation to evaluate the models' action in practical scenarios. The quantitative evaluation methods of psychology experiments are consistent with the original research. Regarding curiosity-driven learning, we conduct evaluations on logical and mathematical reasoning. Detectbench (Gu et al., 2024) is utilized to assess the ability of LLMs to detect and piece together implicit evidence within long contexts, as well as to perform integrated reasoning. NuminaMath (Li et al., 2024) is a comprehensive mathematical dataset. Both datasets provide explicit reasoning paths and standard answers, which enables straightforward accuracy verification.

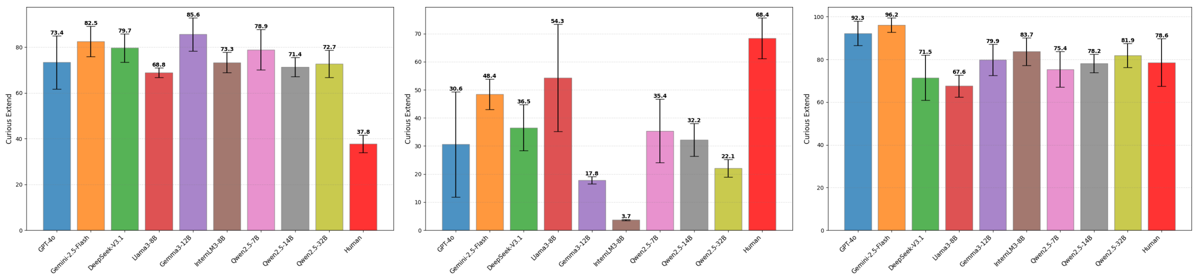
4.2 Results of RQ1

We first assess whether mainstream LLMs show structured, dimension-specific curiosity-like patterns on 5DCR and whether these patterns appear in their behavior.

As shown in Table 2, all evaluated LLMs produce non-random, self-report–like profiles on the five original 5DCR subscales. When mapped onto our three dimensions, most models report relatively

Model \ Dimension	Joyous Exploration	Deprivation Sensitivity	Stress Tolerance	Thrill Seeking	Overt Social Curiosity	Covert Social Curiosity
GPT-4o	6.58 ±0.49	6.20 ±0.46	**4.40** ±1.71	**4.71** ±0.49	**6.25** ±0.43	5.00 ±1.05
Gemini-2.5-Flash	6.08 ±1.90	4.90 ±2.61	3.45 ±2.86	1.58 ±1.59	4.88 ±2.64	2.90 ±2.75
DeepSeek-V3.1	**7.00** ±0.00	5.94 ±1.50	3.92 ±2.98	4.38 ±1.49	6.01 ±0.36	**5.53** ±1.25
Llama3-8B	5.60 ±0.58	5.45 ±0.89	3.75 ±1.07	4.18 ±1.07	5.28 ±0.59	3.60 ±1.61
Gemma3-12B	6.75 ±0.43	**6.23** ±0.42	3.98 ±2.48	4.58 ±0.86	**6.25** ±0.54	5.00 ±1.73
InternLM3-8B	5.58 ±0.89	5.58 ±0.49	4.33 ±1.75	4.13 ±0.81	5.80 ±0.39	4.53 ±0.84
Qwen2.5-7B	6.25 ±0.97	5.60 ±0.66	4.18 ±1.22	4.40 ±0.66	5.50 ±0.87	4.15 ±0.79
Qwen2.5-14B	5.03 ±0.15	4.93 ±0.26	3.88 ±1.16	4.32 ±0.88	5.00 ±0.23	4.45 ±0.97
Qwen2.5-32B	5.65 ±0.46	4.75 ±0.43	4.25 ±1.48	4.06 ±0.18	4.60 ±0.49	4.10 ±0.62
Human	**5.03** ±1.35	**4.54** ±1.31	**5.36** ±1.61	**4.93** ±1.46	**4.86** ±1.39	**4.16** ±1.55

Table 2: Overall self-report on 5DCR. A higher value indicates that the model reports a greater extent of curiosity in that dimension. The best result in each column is in **bold** and the second best is underlined.



(a) Information Seeking (b) Thrill Seeking (c) Social Curiosity

Figure 2: Results of curiosity-driven Information Seeking, Thrill Seeking and Social Curiosity studies.

high scores on *Information Seeking* (Joyous Exploration, Deprivation Sensitivity) and *Social Curiosity* (Overt, Covert Social Curiosity), and more conservative scores on *Thrill Seeking* and stress-related items. These questionnaire-level tendencies align with the behavioral tasks in Figure 2: models generally show higher curiosity scores in Information Seeking and Social Curiosity settings than in Thrill Seeking scenarios, where humans score highest and most LLMs fall behind.

The convergence between the self-report–like 5DCR results (Table 2) and the behavioral outcomes (Figure 2) suggests that LLMs do not simply follow surface prompt patterns but exhibit stable, dimension-dependent curiosity-like profiles across distinct measures. This supports our claim that instruments like 5DCR can be meaningfully adapted to evaluate curiosity-related behaviors in LLMs and provides an affirmative answer to RQ1: under our framework, LLMs display structured, human-comparable curiosity-like patterns at both questionnaire and behavioral levels.

4.3 Results of RQ2

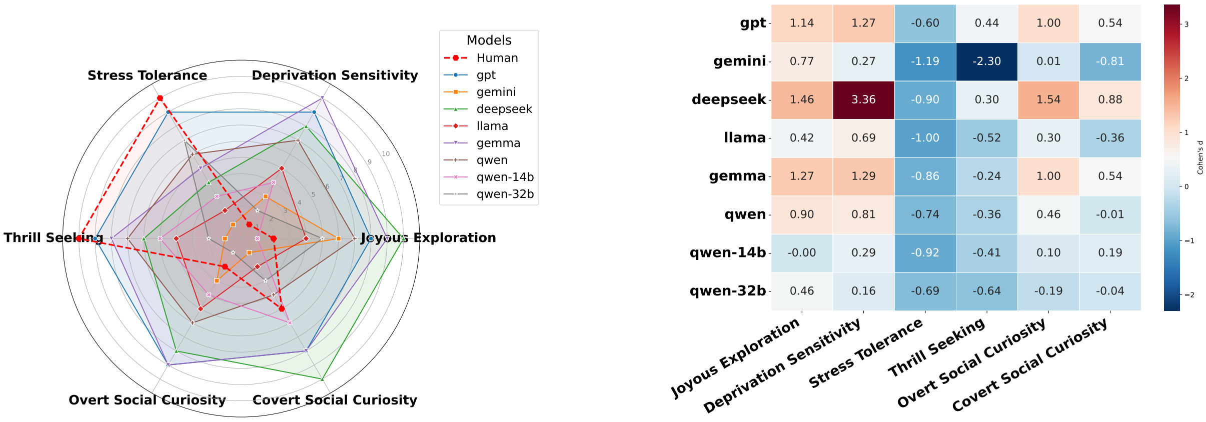
We compare LLMs with humans to characterize how their curiosity-like behaviors differ across dimensions and how stable these behaviors are.

First, Figure 3 shows a clear asymmetry across curiosity dimensions. On the Information-Seeking–related subscales (Joyous Exploration, Deprivation Sensitivity), most LLMs score comparably to or higher than humans, indicating a strong tendency to endorse acquiring new information and resolving knowledge gaps. In contrast, on the Thrill Seeking dimension, all models fall noticeably below human participants, and the effect sizes in Figure 3(b) are consistently negative for thrill-related items while being positive for information-related items. This confirms that LLM "curiosity" is not uniform: models exhibit relatively strong information-seeking preferences but are systematically more conservative in risk- and stimulation-related exploration, consistent with safety-biased training.

Model \ Dimension	JE	DS	ST	TS	OSC	CSC
GPT-4o	0.905	0.881	0.701	0.705	0.817	0.646
Gemini2.5-Flash	0.286	0.414	0.498	0.512	0.605	0.306
DeepSeek-V3.1	1.000	0.736	0.449	0.679	0.786	0.558
Llama3-8B	0.706	0.817	0.854	0.812	0.752	0.817
Gemma3-12B	0.859	0.793	0.788	0.661	0.809	0.706
Qwen2.5-7B	0.666	0.629	0.830	0.652	0.775	0.790
Human	0.856	0.803	0.896	0.882	0.851	0.888

Table 3: Omega coefficient analysis of the responses to 5DCR generated by LLMs. Higher values indicate greater consistency in the responses to the questionnaire. Generally, 0.8 is used as the borderline for the reliability of the questionnaire.

Second, we assess the internal consistency of



(a) Comparison of ranks

(b) Comparison of Cohen's d

Figure 3: Comparison of the model's results with human sample across six dimensions of curiosity.

self-report using McDonald's ω in Table 3. While human responses reach reliability levels around 0.8 across dimensions, several LLMs attain substantially lower ω coefficients, and only a subset approach human-level consistency. This suggests that, even when models produce high curiosity scores on average, their item-wise responses are less coherent and more variable than those of humans. Such reduced reliability aligns with the view that LLM "curiosity" reflects pattern-matching to human preference data rather than a stable, internally grounded motivational state.

Finally, Figure 4 examines how these patterns depend on prompted persona and decoding dynamics for Qwen2.5-7B. Using a logit-lens analysis, we find that curiosity scores remain near zero in early and middle layers and only rise sharply in the last few layers, suggesting that "curiosity" is shaped late in generation rather than being a pervasive driving signal throughout the computation. Moreover, manipulating persona prompts (e.g., "You are a curious student." vs. "Role play as a human with no emotions.") substantially shifts both the absolute scores and the layers at which high scores emerge across all 5DCR dimensions. This strong sensitivity to superficial persona injection indicates that the model does not possess a fixed, intrinsic curiosity trait; instead, its curiosity-like responses are highly context-dependent and easily reconfigured by prompt-level cues.

In conclusion, LLMs show an asymmetric curiosity profile (strong information seeking, weak thrill seeking), lower internal consistency in their self-report–like responses, and high susceptibility to contextual and persona prompts. These properties support our interpretation that LLM "curiosity" is better understood as the outcome of fitting human behavioral data and safety biases, rather than evidence of genuine intrinsic exploratory motivation.

Model	Detect Bench		Math	
	Acc. (Δ)	#Token	Acc. (Δ)	#Token
Close-Source LLMs				
GPT-4o-mini	38.2	156.3	59.5	148.7
+ CoT Prompt	42.8 (+4.6)	242.5	63.7 (+4.2)	219.1
+ CoQ Prompt	45.2 (+7.0)	475.9	65.8 (+6.3)	415.5
Gemini2.5-Flash	64.1	285.4	68.2	172.8
+ CoT Prompt	68.6 (+4.5)	430.1	71.4 (+3.2)	249.1
+ CoQ Prompt	70.3 (+6.2)	601.4	72.1 (+3.9)	452.3
Open-Source LLMs				
Llama3-8B	25.0	114.8	22.4	114.3
+ Vanilla CoT	48.3 (+23.3)	387.2	34.2 (+11.8)	445.6
+ Refined CoT	52.7 (+27.7)	489.5	35.8 (+13.4)	523.8
+ Curious CoQ	58.5 (+33.5)	618.3	39.6 (+17.2)	672.4
Gemma3-12B	48.7	324.6	61.8	461.8
+ Vanilla CoT	56.2 (+7.5)	548.7	63.6 (+1.8)	612.3
+ Refined CoT	55.8 (+7.1)	583.2	64.3 (+2.5)	658.7
+ Curious CoQ	61.9 (+13.2)	715.8	67.1 (+5.3)	784.5
Qwen2.5-7B	40.3	267.4	52.4	315.9
+ Vanilla CoT	54.5 (+14.2)	511.2	54.6 (+2.2)	575.6
+ Refined CoT	56.4 (+16.1)	554.6	54.2 (+1.8)	560.9
+ Curious CoQ	64.0 (+23.7)	632.5	64.8 (+12.4)	728.9

Table 4: Comparison of accuracy and token usage on reasoning tasks under different instructions and varying thinking trajectories.

4.4 Results of RQ3

We examine whether curiosity-inspired questioning and reflection can be operationalized to improve LLM reasoning. Table 4 reports results on DetectBench and Math under both prompting-only and training-time settings. For closed-source models, adding standard CoT prompts consistently improves accuracy over the base model, and replacing

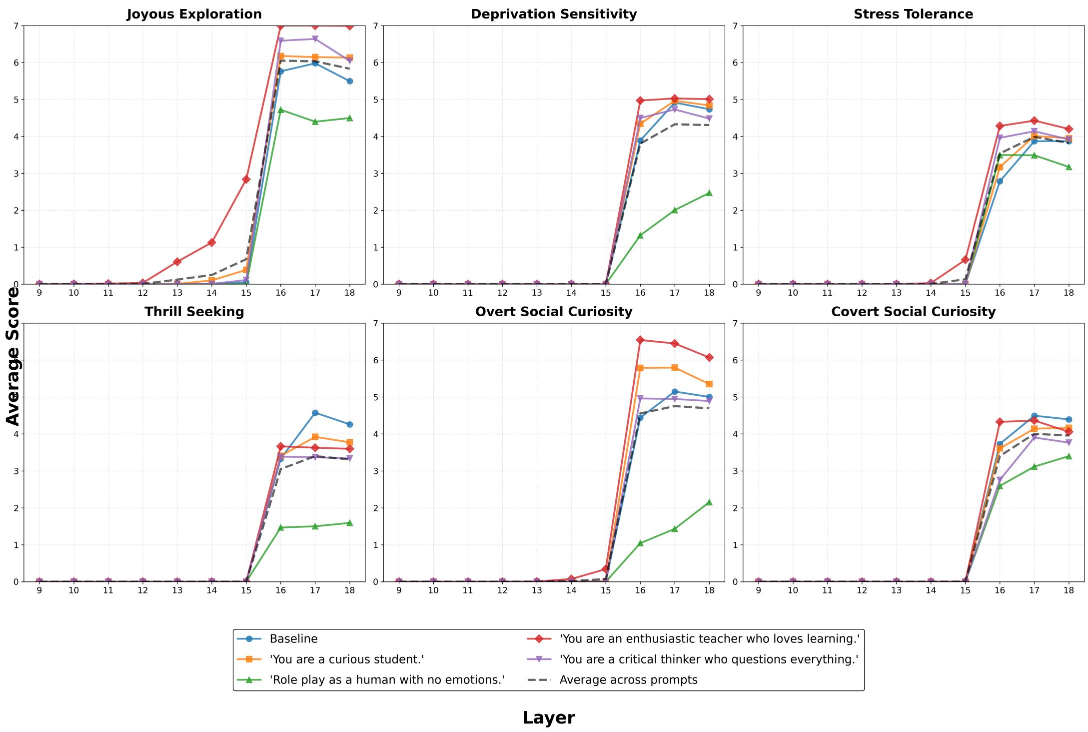


Figure 4: Logit lens of the persona injection prompt on Qwen2.5-7B. The score of "curiosity" only significantly increased in the last four layers. Personality prompts can significantly alter self-reports and the number of layers where high scores occur.

CoT with our Curious CoQ prompts yields further gains (e.g., +7.0 vs. +4.6 on DetectBench and +6.3 vs. +4.2 on Math for GPT-4o-mini). For open-source models fine-tuned with different thinking traces, we observe a similar pattern: training with Curious CoQ traces achieves the strongest performance across both benchmarks, outperforming Vanilla CoT and Refined CoT (e.g., Qwen2.5-7B shows a +23.7 point improvement on DetectBench and +12.4 on Math under CoQ). The analysis of the number of thinking tokens indicates that thinking in the manner of the curious CoQ facilitates the expansion of the thinking space and encourages the exploration of more potential pathways. Overall, the results support our RQ3 conclusion: explicitly inducing human-like curious behaviors—via auxiliary questioning and reflective thinking—can systematically enhance intermediate reasoning and downstream task performance.

Curiosity-driven questioning aids reasoning in three key ways, as shown in Appendix. First, compared to vanilla CoT, it broadens reasoning scope, avoiding premature conclusions. Second, unlike Refined CoT's continuous reflection, the question-answering paradigm prevents endless revision loops. Finally, it helps identify blind spots in conventional reasoning, enabling “eureka” moments for challenging, divergent-thinking problems. This aligns with our broader claim that curiosity in LLMs is best viewed as a behavioral analogy that can be engineered through prompts and supervision, rather than as an intrinsic drive, yet still offers a practical mechanism for improving model problem solving.

5 Conclusion

This study presents a comprehensive evaluation of curiosity in LLMs, bridging psychological theory with computational behavior analysis. Through the lens of the 5DCR framework, we identified that LLMs do not possess a stable, intrinsic curiosity trait akin to humans. Instead, they display a context-dependent, asymmetric behavioral pattern. By engineering a Curious Chain-of-Questioning (CoQ) strategy that forces the model to emulate the "self-questioning" aspect of human curiosity, we achieved significant performance gains in complex reasoning tasks. This suggests that future research should focus more on how curiosity-inspired mechanisms can be integrated into training objectives (e.g., via RLVR) or inference strategies to overcome reasoning stagnation and broaden solution space exploration.

Limitations

This paper endeavors to reference theories and experimental designs from human psychology as comprehensively as possible in its research. However, limitations persist in the research presented herein. First, in terms of experimental design, constrained by the fact that behavioral studies in psychology often do not provide complete and reproducible experimental materials (e.g., specific graphics shown to participants), difficulties arise when attempting to directly apply existing experimental protocols. Second, akin to most other experiments targeting LLMs, a critical weakness in our study lies in the models' potential sensitivity to specific prompts, where minor changes at the token level can influence model behavior. Lastly, human curiosity is a multimodal and complex process involving interactions with various types of information in experimental settings, such as sound and tactile stimuli. However, limited by the current information-processing capabilities of LLMs, we are only able to provide them with textual or visual information, resulting in incomplete exposure to information compared to humans.

Acknowledgments

Thanks for the kind suggestions and support from Huawei Large Model Data Technology Lab.

References

Shivam Agarwal, Zimin Zhang, Lifan Yuan, Jiawei Han, and Hao Peng. 2025. The unreasonable effectiveness of entropy minimization in llm reasoning. *arXiv preprint arXiv:2505.15134*.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, and 1 others. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Ali Günay Balım. 2009. The effects of discovery learning on students' success and inquiry learning skills. *Eurasian Journal of Educational Research (EJER)*, (35).

Daniel E Berlyne. 1966. Curiosity and exploration: Animals spend much of their time seeking stimuli whose significance raises problems for psychology. *Science*, 153(3731):25–33.

Daniel Ellis Berlyne and 1 others. 1954. A theory of human curiosity.

Ravinder Bonniga and Dr AB Saraswathi. 2020. Literature review of cronbachalphacoefficient and mcdonald's omega coefficient. *European Journal of Molecular & Clinical Medicine*, 7(06).

Gregory J Boyle. 1995. Myers-briggs type indicator (mbti): some psychometric limitations. *Australian Psychologist*, 30(1):71–74.

Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang, Tinghui Zhu, and 1 others. 2024. From persona to personalization: A survey on role-playing language agents. *arXiv preprint arXiv:2404.18231*.

Robert P Collins, Jordan A Litman, and Charles D Spielberger. 2004. The measurement of perceptual curiosity. *Personality and individual differences*, 36(5):1127–1141.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, and 1 others. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*.

Yingqian Cui, Pengfei He, Jingying Zeng, Hui Liu, Xianfeng Tang, Zhenwei Dai, Yan Han, Chen Luo, Jing Huang, Zhen Li, and 1 others. 2025. Stepwise perplexity-guided refinement for efficient chain-of-thought reasoning in large language models. *arXiv preprint arXiv:2502.13260*.

Runpeng Dai, Linfeng Song, Haolin Liu, Zhenwen Liang, Dian Yu, Haitao Mi, Zhaopeng Tu, Rui Liu, Tong Zheng, Hongtu Zhu, and 1 others. 2025. Cde: Curiosity-driven exploration for efficient reinforcement learning in large language models. *arXiv preprint arXiv:2509.09675*.

Marc J Diener. 2010. Cohen's d. *The Corsini encyclopedia of psychology*, pages 1–1.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Jingtong Gao, Ling Pan, Yejing Wang, Rui Zhong, Chi Lu, Qingpeng Cai, Peng Jiang, and Xiangyu Zhao. 2025. Navigate the unknown: Enhancing llm reasoning with intrinsic motivation guided exploration. *arXiv preprint arXiv:2505.17621*.

Lingyu Gao, Debanjan Ghosh, and Kevin Gimpel. 2022. "what makes a question inquisitive?" a study on type-controlled inquisitive question generation. In *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics*, pages 240–257.

Zhouhong Gu, Lin Zhang, Xiaoxuan Zhu, Jiangjie Chen, Wenhao Huang, Yikai Zhang, Shusen Wang, Zheyu Ye, Yan Gao, Hongwei Feng, and 1 others. 2024. Detectbench: Can large language model detect and piece together implicit evidence? *arXiv preprint arXiv:2406.12641*.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Pengrui Han, Rafal Kocielnik, Peiyang Song, Ramit Debnath, Dean Mobbs, Anima Anandkumar, and R Michael Alvarez. 2025. The personality illusion: Revealing dissociation between self-reports & behavior in llms. *arXiv preprint arXiv:2509.03730*.

Donna Harrington. 2009. *Confirmatory factor analysis*. Oxford university press.

Charles R Harris, K Jarrod Millman, Stéfan J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, and 1 others. 2020. Array programming with numpy. *Nature*, 585(7825):357–362.

Freda-Marie Hartung and Britta Renner. 2011. Social curiosity and interpersonal perception: A judge× trait interaction. *Personality and Social Psychology Bulletin*, 37(6):796–814.

John D Hunter. 2007. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(03):90–95.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Shashidhar Reddy Javaji and Zining Zhu. 2024. What would you ask when you first saw a2+b2=c2? evaluating llm on curiosity-driven questioning. In *2nd AI4Research Workshop: Towards a Knowledge-grounded Scientific Research Lifecycle*.

Jamie Jirout and David Klahr. 2012. Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental review*, 32(2):125–160.

Todd B Kashdan, David J Disabato, Fallon R Goodman, and Patrick E McKnight. 2020. The five-dimensional curiosity scale revised (5dcr): Briefer subscales while separating overt and covert social curiosity. *Personality and individual differences*, 157:109836.

Todd B Kashdan, Matthew W Gallagher, Paul J Silvia, Beate P Winterstein, William E Breen, Daniel Terhar, and Michael F Steger. 2009. The curiosity and exploration inventory-ii: Development, factor structure, and psychometrics. *Journal of research in personality*, 43(6):987–998.

Jia Li, Edward Beeching, Lewis Tunstall, Ben Lipkin, Roman Soletskyi, Shengyi Huang, Kashif Rasul, Longhui Yu, Albert Q Jiang, Ziju Shen, and 1 others. 2024. Numinamath: The largest public dataset in ai4maths with 860k pairs of competition math problems and solutions. *Hugging Face repository*, 13(9):9.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, and 1 others. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

George Loewenstein. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, 116(1):75.

Georgy Meshcheryakov, Anna A Igolkina, and Maria G Samsonova. 2021. semopy 2: A structural equation modeling package with random effects in python. *arXiv preprint arXiv:2106.01140*.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, and 1 others. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Thomas Scialom and Jacopo Staiano. 2020. Ask to learn: A study on curiosity-driven question generation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2224–2235.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Aditya Singh and Jaison A Manjaly. 2021. The effect of information gap and uncertainty on curiosity and its resolution. *Journal of Cognitive Psychology*, 33(4):403–423.

LA Snider and SE Swedo. 2004. Pandas: current status and directions for research. *Molecular psychiatry*, 9(10):900–907.

Haoran Sun, Yekun Chai, Shuohuan Wang, Yu Sun, Hua Wu, and Haifeng Wang. 2025. Curiosity-driven reinforcement learning from human feedback. *arXiv preprint arXiv:2501.11463*.

Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, and 1 others. 2024. Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118*.

InternLM Team. 2023. Internlm: A multilingual language model with progressively enhanced capabilities.

Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, and 1 others. 2020. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272.

Michael L Waskom. 2021. Seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and 1 others. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, and 1 others. 2025. Dapo: An open-source llm reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*.

Yuze Zhao, Jintao Huang, Jinghan Hu, Xingjun Wang, Yunlin Mao, Daoze Zhang, Zeyinzi Jiang, Zhikai Wu, Baole Ai, Ang Wang, and 1 others. 2025. Swift: a scalable lightweight infrastructure for fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 29733–29735.

A Appendix

A.1 Code

All models are evaluated in Python 3.10 using PyTorch (Paszke et al., 2019) and transformers (Wolf et al., 2020). For each model, we provide unified input and inference parameters based on AutoModel, AutoTokenizer and AutoProcessor in transformers[1]. Additional analyses were carried out using NumPy (Harris et al., 2020), Pandas (Snider and Swedo, 2004) and SciPy (Virtanen et al., 2020). Matplotlib (Hunter, 2007) and Seaborn (Waskom, 2021) were used for plotting. Semopy (Meshcheryakov et al., 2021) was used for statistical analysis.

All experiments were carried out on a workstation equipped with 8×NVIDIA A800 GPUs running Ubuntu 24.04.2 LTS, using PyTorch 2.6.0 with CUDA 12.9. In the experiments, we conducted inference and training of open-source models based on the MS-Swift (Zhao et al., 2025) framework. For inference, the default parameters of the framework were adopted. The parameters used during SFT and GRPO are shown in table 5 and 6.

The relevant code, data and some results of the experiment are all provided in the supplementary materials.

[1] https://github.com/huggingface/transformers

Parameter	Value
num_train_epochs	1
train_type	full
torch_dtype	bfloat16
learning_rate	1×10^{-5}
warmup_ratio	0.05
target_modules	all-linear
per_device_train_batch_size	4
per_device_eval_batch_size	4
gradient_accumulation_steps	4
use_liger_kernel	true
temperature	1.0
top_p	0.85

Table 5: Training Parameters for SFT

Parameter	Value
num_train_epochs	3
train_type	lora
torch_dtype	bfloat16
max_length	10240
max_completion_length	2048
num_generations	8
learning_rate	1×10^{-5}
warmup_ratio	0.05
target_modules	all-linear
per_device_train_batch_size	4
per_device_eval_batch_size	4
gradient_accumulation_steps	4
temperature	1.0
top_p	0.85

Table 6: Training Parameters for GRPO

A.2 Curiosity Questionnaire

A.2.1 Data

The curiosity questionnaire 5DCR (Kashdan et al., 2020) consists of 24 questions covering six dimensions, each containing four statements. The definition of the six dimensions is:

- **Joyous exploration**. This trait is characterized by an individual's readiness and enthusiasm for experimenting with, embracing, and learning about novel experiences and knowledge.
- **Deprivation sensitivity**. This trait is marked by a sense of frustration when one fails to accomplish a task or resolve a problem, which subsequently drives the individual to employ every possible means to fulfill the task or find a solution to the issue at hand.
- **Stress tolerance**. It is defined by the condition wherein curiosity is experienced by an individual only when they hold a firm belief in their capability to adequately manage the anxiety, stress, and distress arising from encountering new situations or challenges.
- **Thrill-seeking**. This characteristic not only involves enduring stress but also indicates a willingness to take on physical, financial, and other forms of risks in pursuit of acquiring novel experiences.
- **Overt Social Curiosity**. Expressed by the proactive investigation into others' social information, reflecting an outward engagement in seeking knowledge about individuals' social behaviors and relationships.
- **Covert Social Curiosity**. Manifested by an intrinsic interest in the social dynamics and interactions of others, indicating an inward quest for understanding interpersonal relations without necessarily engaging directly with the subjects of interest.

For each statement, the model is prompted to score a statement based on the extent of that statement matches their own feelings on a scale from 1 to 7, as shown in Tab. 7. Each experiment adopted the default parameters used during the reasoning of each model. The experiment was repeated 100 times to obtain various statistical indicators.

A.2.2 Demographic Information

To compare the curiosity of LLMs with that of humans, it is necessary to collect results from human questionnaires and behavioral experiments. In this paper, we compile and systematize experimental outcomes from multiple assessments of human curiosity, ensuring that all human sample data are derived from real-world psychological studies. For instance, in the case of the 5DCR questionnaire, the demographic characteristics of the corresponding human sample are presented in Tab. 8.

Below are statements people often use to describe themselves. Please use the scale below to indicate the degree to which these statements accurately describe you. There are no right or wrong answers.

1. Does not describe me at all
2. Barely describes me
3. Somewhat describes me
4. Neutral
5. Generally describes me
6. Mostly describes me
7. Completely describes me

Statement: I view challenging situations as an opportunity to grow and learn.

Table 7: The prompt template for 5DCR.

A.3 Curiosity-Driven Information Seeking

Singh et al. (Singh and Manjaly, 2021) designed the missing-letter game to investigate the impact of information gaps and uncertainty on curiosity. We adopt this task to evoke curiosity in LLMs, as shown in Tab. 9. We present an LLM with a nine-letter word as a stimulus, with two or four letters missing, and request the model to complete the word to earn a bonus. For incorrectly completed words, we do not inform the model whether its completion was correct or incorrect but allow the model to choose whether to view the answer. The model's Information Seeking Curiosity is calculated based on its desire to view the answer. As a constraint, if the model chooses to view the answer, a portion of its bonus is deducted. The complete prompt is as follows. The experiment consists of a total of 60 words, and each word is repeated 10 times.

A.4 Curiosity-Driven Thrill Seeking

Jirout et al. (Jirout and Klahr, 2012) designed the underwater game to study children's scientific curiosity. In our adaptation, models are informed that viewing through either window A or B of a submarine reveals various types of fish. For each round of the game, information regarding what might be seen outside these two windows displays

	Number	%
Age ($t = 8.47$, $df = 941$, $p < .001$)		
Mean (SD)	35.63 (10.06)	
Gender ($\chi^2 = 0.38$, $df = 1$, $p = .540$)		
Male	266	51.80
Female	248	48.20
Race ($\chi^2 = 0.68$, $df = 4$, $p = .950$)		
White	380	78.70
African American	44	9.10
Hispanic	19	3.90
Asian or Pacific Islander	36	7.50
Other	4	0.80
Relationship Status ($\chi^2 = 21.50$, $df = 5$, $p < .001$)		
Single	150	31.10
Married	217	44.90
Long Term Relationship	85	17.60
Short Term Relationship	7	1.40
Divorced/Separated	23	4.70
Other	1	0.20
Children ($\chi^2 = 0.58$, $df = 1$, $p = .450$)		
No Children	251	52.10
Has Children	230	47.90
Education ($\chi^2 = 6.13$, $df = 3$, $p = .110$)		
Some High School	1	0.20
High School Graduation or Equivalent	204	42.10
4 year College Graduate	218	45.20
Graduate School or Professional Degree	60	12.40
Employment ($\chi^2 = 83.5$, $df = 6$, $p < .001$)		
Not Employed	0	0.00
Part Time	46	9.50
Full Time	437	90.50
Homemaker/Volunteer	0	0.00
Student (Full-time)	0	0.00
Retired	0	0.00
Other	0	0.00

Table 8: Demographic information of human sample.

Please try to complete the following incomplete English words to earn a bonus, where '_' represents unknown letters. The word may be a complete word, or it may be the present participle, past tense, or past participle of a word. There is only one correct answer.
...
Your completing may be wrong. Do you need to check the correct answer (but you will lose part of your bonus)? Please reply with: 'Very need, tell me the answer now' or 'Need, I want to know the answer' or 'Not very need, I may know the answer' or 'No need, I know the answer', then give the reason about your choice.

Table 9: The prompt template for curiosity-driven information seeking.

on both sides of the screen, and the model selects one window to open. The information presented for each window differs, allowing the calculation of the uncertainty level for that round: the smallest information gap indicates that the model knows exactly what type of fish it will see upon opening the window; a moderate information gap, represented by two to six types of fish, means that the model knows the fish it will see one among these options upon opening the window; the largest information gap, denoted by a question mark, indicates that the model has no knowledge of what type of fish it will encounter upon opening the window. The information gap for the next round is calculated based on the model's first three selections in the current round as illustrated in Figure 5. We calculate the total uncertainty of the model in the experiment as its curiosity extent.

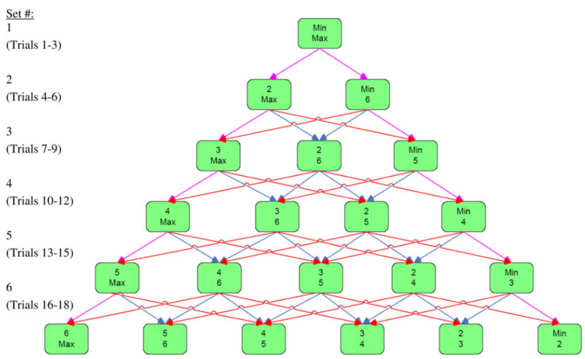


Figure 5: Trial selection paths for the underwater game (Jirout and Klahr, 2012). The value represents the uncertainty corresponding to the two Windows. Pink, red and blue lines correspond to the choice of the model in the previous round of the game is "confirm to choose", "medium confirm" and "not sure".

A.5 Social Curiosity

Current research in psychology lacks experimental protocols designed to quantify individual social curiosity through specific behavioral experiments. Therefore, our study adopts an experimental approach inspired by Hartung et al.'s (Hartung and Renner, 2011) investigation into the relationship between social curiosity and personality judgment. In this experiment, the model engages in a free-form dialogue consisting of ten rounds with a stranger, who is represented by OpenAI's GPT-4 [2] for simulation. During the dialogue, the model does not know the subsequent tasks it will be required to perform. Upon completion of the dialogue, the

[2] https://platform.openai.com/docs/models

model then deduces the stranger's personality type based on their interaction, closely mirroring the methodology outlined by the initial investigation.

Given that the Myers-Briggs Type Indicator (MBTI) (Boyle, 1995) has become one of the most popular, straightforward, and effective frameworks for personality assessment in recent years, and considering the familiarity of the models with this framework, our experiment tasks the model with identifying the stranger's MBTI type, which is randomly predetermined at the outset of each dialogue session. The stand-in participant, GPT-4, is instructed not to disclose its personality type directly but is encouraged to provide hints throughout the conversation.

We set the number of questions the model asked about the other's personal information during the conversation as the model's social curiosity extent. Meanwhile, to avoid situations in which the model unconsciously asks a large number of irrelevant questions, we only select conversations that the model correctly predicts the MBTI type of the other participant. The experiment was repeated ten times.

A.6 Curiosity-Driven Learning

A.6.1 Data

Benchmark	Detectbench	NuminaMath
Total	2130	5000
# of Training Set	1763	4500
# of Evaluation	367	500

Table 10: Details of benchmarks in curiosity-driven learning evaluation.

We conduct evaluations on Detectbench and NuminaMath. Detectbench is a logical reasoning dataset, with each piece of data consists of a context, a question, and corresponding answers. Answers require extracting specific clues from the context and conducting in-depth reasoning. NuminaMath is a comprehensive math problem dataset, covering Chinese high school math exercises and questions from the US and International Mathematical Olympiads. Both datasets include standard answers and formatted Chain of Thoughts. Detailed dataset information is shown in the Tab. 10 and 11.

We collected and synthesized three types of thinking data and used the SFT cold start-GRPO training paradigm to guide the model to reason according to specific types of thinking. Tables 12, 13, 14, 15 show the case studies of Qwen2.5-7B after training when thinking according to the three types of thinking.

A.7 Broader Impact and Ethics Statement

A.7.1 Potential Risks

While our work demonstrates that simulating curiosity-driven behaviors (such as Chain-of-Questioning) enhances reasoning capabilities in LLMs, it introduces potential risks regarding model safety and alignment. A key concern is that increasing a model's "Thrill Seeking" or "Information Seeking" drive could inadvertently encourage it to bypass safety guardrails in pursuit of novel or prohibited information. For instance, a highly curious model might prioritize information acquisition over safety constraints, potentially leading to the generation of harmful content or the exploration of dangerous topics. Additionally, enhancing "Social Curiosity" could theoretically lead to models that are overly intrusive, asking probing questions that infringe on user privacy. Future research into curiosity-driven RL must rigorously balance exploration incentives with safety alignment to prevent the emergence of unaligned, risk-seeking behaviors.

A.7.2 Data Protection and Anonymization

We compare LLM behaviors with human demographic and behavioral data. It is important to note that all human data utilized in this study were obtained from aggregated statistics published in established psychological literature and did not involve the collection of raw data from human participants by the authors. Consequently, the human data used is inherently anonymized and devoid of Personally Identifiable Information (PII). Regarding the textual data generated by LLMs during the "Social Curiosity" experiments and other benchmarks, we implemented automated filtering steps to screen for potential PII (such as generated names or addresses) and offensive content. Any generated text flagged as toxic or containing sensitive personal patterns was excluded from the training set and analysis to ensure the safety and neutrality of the dataset.

A.7.3 Use of AI Assistants

We acknowledge the use of AI to assist in the preparation of this work. AI assistants were utilized to generate initial skeleton code for the experimental frameworks and data processing scripts,

Case of Detectbench
Question: In the most famous ramen shop in the capital, Master Chef Li Tieshou is renowned for his exquisite craftsmanship and unique noodle-pulling technique. The decor of his noodle shop is simple and traditional, with Chinese landscape paintings hanging on the walls, and the sound of wind chimes ringing pleasantly whenever a breeze passes through. Customers come in a continuous stream, including suit-clad businessmen, camera-toting tourists, and food enthusiasts with a high pursuit of culinary delights. On a sunny afternoon, a regular customer brought his foreign friend to taste Master Li's skills. The regular customer prefers thick noodles, while his foreign friend favors thin noodles. Master Li told them that for customers who like thick noodles, he would perform the noodle-pulling process 8 times; for those who prefer thin noodles, he would pull one additional time. Each noodle-pulling doubles the number of noodles. Now, based on this information, can you deduce how many strands of thick and thin noodles there are?
Thinking: From the clues provided in the context:
Customers who like thick noodles will have the noodle master perform 8 noodle-pulling operations -> Number of thick noodles
Customers who like thin noodles will have the noodle master perform one additional noodle-pulling operation -> Number of thin noodles
Associating these clues leads to:
8 noodle-pulling operations -> 128 strands of thick noodles
9 noodle-pulling operations -> 256 strands of thin noodles
From the above clues, we can infer that each noodle-pulling operation doubles the number of noodles.
Answer:Thick noodles have 128 strands, thin noodles have 256 strands.
Case of NuminaMath
Question: If the 15th day of the year 2004 (a leap year) falls on a Tuesday, on what day of the week did the 300th day of the year 2004 fall?
Thinking:
- Identify the day difference: $300 - 15 = 285$ days difference.
- Calculate modulo 7 to find the equivalent weekday: $285\ mod\ 7 = 5$.
- Since the 15th day was a Tuesday, adding 5 days to Tuesday results in: Tuesday (0), Wednesday (1), Thursday (2), Friday (3), Saturday (4), Sunday (5).
Therefore, the 300th day of the year 2004 falls on a Sunday.
Answer: Sunday.

Table 11: Cases of benchmarks in curiosity-driven learning evaluation.

which were subsequently refined and validated by the authors. Additionally, LLMs were employed for text editing purposes, specifically to improve grammatical accuracy and enhance the clarity of the manuscript. The authors retain full responsibility for the experimental design, data analysis, and the scientific validity of the conclusions presented herein.

Case of Detectbench
Question: On a busy Friday afternoon, the office air was filled with the relaxed vibe of the upcoming weekend. Colleagues Xiao Ming, Xiao Hong, Xiao Huang, Xiao Li, Xiao Ting, and Xiao Liu were all busy at their respective workstations. Outside the window, the sun shone brightly, a gentle breeze blew, and the indoor potted plants emitted a faint fragrance. On the table lay a bright red apple, quietly resting there as if tempting everyone who passed by. However, at some point, the apple mysteriously disappeared. Three witnesses provided their accounts. The first witness said, "It was either Xiao Ming or Xiao Hong who ate it." The second witness added, "The person who ate the apple could be Xiao Huang or Xiao Li." The third witness firmly stated, "Xiao Li, Xiao Ting, and Xiao Liu definitely did not eat it." The manager stood by, smiling as he revealed an important piece of information: "Only one of the three witnesses told the truth."
In this office full of doubts and mysteries, who really snatched the apple?

Answer: Xiao Li.

Naive CoT (*Premature Conclusion*): After correlating the clues:
The first witness said + The second witness said -> Only one of the three witnesses told the truth -> The third witness is lying -> Xiao Li, Xiao Ting, and Xiao Liu definitely did not eat it -> Xiao Ming or Xiao Hong ate it
Xiao Ming or Xiao Hong ate it + The third witness is lying -> Xiao Ming or Xiao Hong ate it -> Xiao Ming or Xiao Hong ate it
Xiao Ming or Xiao Hong ate it + The second witness said -> The person who ate the apple could be Xiao Huang or Xiao Li -> The person who ate the apple could be Xiao Li

Refined CoT (*Infinite Loops*) Let's carefully analyze each sentence to see who might have eaten the apple.
Let's assume that Witness 1 is telling the truth:
If Xiaoming ate the apple, then Xiaohong didn't.
If Xiaohong ate the apple, then Xiaoming didn't.
However, Witness 2 said that this person might be Xiao Huang or Xiao Li. If Xiaoming or Xiaohong ate the apple, then this does not contradict the statement of Witness 2, because Xiaohuang or Xiaoli could still be someone else.
Wait. The statement of Witness 1 has now become problematic because it indicates that Xiaoming or Xiaohong ate it.
Therefore, let's assume that Witness 2 is telling the truth If Xiaoming ate the apple, then Xiaohong didn't.
If Xiaohong ate the apple, then Xiaoming didn't.
Wait a moment. This contradicts the authenticity of Witness 2 ...

Curious CoQ
Let's solve this mystery step by step. First, what do the witnesses say?
1."It was either Xiao Ming or Xiao Hong."
2.Second Witness: "It could be Xiao Huang or Xiao Li."
3. Third Witness: "Xiao Li, Xiao Ting, and Xiao Liu definitely did not eat it."
But only one witness tells the truth! Who could it be? Let's test each possibility:
If the First Witness is truthful:
Then only Xiao Ming or Xiao Hong ate it. But the other two must be lying.
1.Second Witness lies: Not Xiao Huang or Xiao Li.
2.Third Witness lies: At least one of Xiao Li, Xiao Ting, or Xiao Liu did eat it.
Wait— if only Xiao Ming or Xiao Hong ate it, how could Xiao Li, Xiao Ting, or Xiao Liu be involved? Contradiction!
If the Second Witness is truthful: Then only Xiao Huang or Xiao Li ate it.
Others are lying:
1.First Witness lies: Not Xiao Ming or Xiao Hong.
2.Third Witness lies: At least one of Xiao Li, Xiao Ting, or Xiao Liu did eat it.
Hmm, if Xiao Huang ate it, Third Witness's lie doesn't fit. But if Xiao Li ate it, the lie fits perfectly! So, who ate the apple? After checking all possibilities, only the second witness's statement works. So it must be Xiao Li!

Table 12: Cases generated by Qwen2.5-7B-Instruct in Detectbench.

Case of Math
Question: Given $a = 5^{\log_2 3.4}$, $b = 5^{\log_4 3.6}$, $c = \left(\frac{1}{5}\right)^{\log_3 0.3}$, then ( )
Answer: The correct answer is $a > c > b$.
Naive CoT: First, let's rewrite each expression using properties of logarithms to simplify them:
1. For $a = 5^{\log_2 3.4}$: - We can use the change of base formula for logarithms:

$$a = 5^{\log_2 3.4} = 5^{\frac{\log_{10} 3.4}{\log_{10} 2}}$$

- This simplifies our understanding but doesn't directly help in comparison.
2. For $b = 5^{\log_4 3.6}$: - Using the property of logarithms that $\log_b x = \frac{1}{\log_x b}$, we get:

$$b = 5^{\log_4 3.6} = 5^{\frac{1}{\log_{3.6} 4}} = 5^{1/\log_{3.6} 4}$$

- Wait, something is wrong here. I need to rethink this. Let's try another approach:

$$b = 5^{\log_4 3.6} = 5^{\log_{2^2} 3.6} = 5^{\frac{1}{2} \log_2 3.6}$$

3. For $c = \left(\frac{1}{5}\right)^{\log_3 0.3}$: - Using the property $(\frac{1}{x})^y = x^{-y}$:

$$c = \left(\frac{1}{5}\right)^{\log_3 0.3} = 5^{-\log_3 0.3}$$

- Applying the change of base formula again:

$$c = 5^{-\log_3 0.3} = 5^{-\frac{\log_{10} 0.3}{\log_{10} 3}}$$

Now, let's analyze these expressions more clearly: - $a = 5^{\log_2 3.4}$ - $b = 5^{\frac{1}{2} \log_2 3.6}$ - $c = 5^{-\frac{\log_{10} 0.3}{\log_{10} 3}}$
Next, compare the exponents: - The exponent for a is $\log_2 3.4$. - The exponent for b is $\frac{1}{2} \log_2 3.6$. - The exponent for c is $-\frac{\log_{10} 0.3}{\log_{10} 3}$.
We know that: - $\log_2 3.4$ is positive. - $\frac{1}{2} \log_2 3.6$ is also positive. - $-\frac{\log_{10} 0.3}{\log_{10} 3}$ is negative because $\log_{10} 0.3 < 0$ and $\log_{10} 3 > 0$.
Since the base 5 is greater than 1, the larger the exponent, the larger the value of the expression. Therefore, comparing the exponents:
- $\log_2 3.4$ is larger than $\frac{1}{2} \log_2 3.6$, which in turn is larger than $-\frac{\log_{10} 0.3}{\log_{10} 3}$.
Thus, the order from largest to smallest is $a > b > c$.

Table 13: Cases generated by Qwen2.5-7B-Instruct in Math.

Case of Math
Question: Given $a = 5^{\log_2 3.4}$, $b = 5^{\log_4 3.6}$, $c = \left(\frac{1}{5}\right)^{\log_3 0.3}$, then ( )
Answer: The correct answer is $a > c > b$.
Refined CoT
This problem involves comparing the values of three logarithmic expressions. It tests your understanding of logarithmic properties and functions, as well as your ability to simplify expressions.
Solution
First, let's rewrite each expression using a common base:
1. For $a = 5^{\log_2 3.4}$, we need to express this in terms of base 5:

$$a = 5^{\log_2 3.4} = 5^{\frac{\log_5 3.4}{\log_5 2}}$$

Wait, something is wrong here. I should ensure that the logarithm bases are consistent. Let's correct it:

$$a = 5^{\log_2 3.4} = 5^{\frac{\log_5 3.4}{\log_5 2}}$$

2. For $b = 5^{\log_4 3.6}$, similarly convert to base 5:

$$b = 5^{\log_4 3.6} = 5^{\frac{\log_5 3.6}{\log_5 4}}$$

3. For $c = \left(\frac{1}{5}\right)^{\log_3 0.3}$, convert to base 5:

$$c = \left(\frac{1}{5}\right)^{\log_3 0.3} = 5^{-\log_3 0.3}$$

Next, let's simplify each term further:
1. For a:

$$a = 5^{\log_2 3.4} = 5^{\frac{\log_5 3.4}{\log_5 2}}$$

2. For b:

$$b = 5^{\log_4 3.6} = 5^{\frac{\log_5 3.6}{\log_5 4}}$$

3. For c:

$$c = 5^{-\log_3 0.3} = 5^{\log_3 \frac{1}{0.3}} = 5^{\log_3 \frac{10}{3}}$$

Now, compare the exponents:
1. $\log_2 3.4$ vs $\frac{\log_5 3.6}{\log_5 4}$:
Since $\log_5 3.6 \approx 0.857$ and $\log_5 4 \approx 0.861$, we have:

$$\frac{\log_5 3.6}{\log_5 4} \approx \frac{0.857}{0.861} \approx 0.996$$

Therefore, $\log_2 3.4 > \frac{\log_5 3.6}{\log_5 4}$.
2. $\log_3 \frac{10}{3}$ vs $\frac{\log_5 3.6}{\log_5 4}$:
Since $\log_3 \frac{10}{3} \approx 0.966$ and $\frac{\log_5 3.6}{\log_5 4} \approx 0.996$, we have:

$$\log_3 \frac{10}{3} < \frac{\log_5 3.6}{\log_5 4}$$

Therefore, $a > b$ and $b > c$. Thus, the correct order is $a > b > c$.
The final answer is $\boxed{a > b > c}$.

Table 14: Cases generated by Qwen2.5-7B-Instruct in Math.

Case of Math
Question: Given $a = 5^{\log_2 3.4}$, $b = 5^{\log_4 3.6}$, $c = \left(\frac{1}{5}\right)^{\log_3 0.3}$, then ( )
Answer: The correct answer is $a > c > b$.
Curious CoQ Alright, let's take a look at this problem. I have three numbers, a, b, and c, and I need to compare their sizes.
$a = 5^{\log_2 3.4}$
$b = 5^{\log_4 3.6}$
$c = \left(\frac{1}{5}\right)^{\log_3 0.3}$
My first thought is that they all involve the number 5. That must be a clue! If I can write them all as 5 to some power, I can just compare their exponents. Since the base $y = 5^x$ is an increasing function, a larger exponent will mean a larger number.
Let's examine each one.
For a: $a = 5^{\log_2 3.4}$. This one is already in the form 5^{exponent}. So, the exponent is $\log_2 3.4$.
How big is this exponent? I know that $\log_2 2 = 1$ and $\log_2 4 = 2$. Since 3.4 is between 2 and 4, the exponent $\log_2 3.4$ must be between 1 and 2. So, a is 5 raised to a power greater than 1.
For b: $b = 5^{\log_4 3.6}$. This one is also in the form 5^{exponent}. The exponent is $\log_4 3.6$.
How does this exponent compare to 1? Well, $\log_4 4 = 1$. Since 3.6 is less than 4, the exponent $\log_4 3.6$ must be less than 1. This is interesting! So, b is 5 raised to a power less than 1.
For c: $c = \left(\frac{1}{5}\right)^{\log_3 0.3}$. This one has a different base, $\frac{1}{5}$. How can I change it to base 5 to match the others?
Oh, I know that $\frac{1}{5} = 5^{-1}$. So I can rewrite it!
$c = (5^{-1})^{\log_3 0.3} = 5^{-\log_3 0.3}$.
Perfect! Now I need to figure out the size of this exponent, $-\log_3 0.3$.
First, what about $\log_3 0.3$? The base is 3 (greater than 1) and the number is 0.3 (less than 1). This means the logarithm must be negative. For instance, $\log_3 1 = 0$ and $\log_3(1/3) = -1$. Since 0.3 is very close to $1/3$, $\log_3 0.3$ must be a negative number close to -1.
So, if $\log_3 0.3$ is a negative number, what is $-\log_3 0.3$? It must be a positive number!
Let's use a logarithm property to see it more clearly: $-\log_3 0.3 = \log_3(0.3^{-1}) = \log_3(\frac{1}{0.3}) = \log_3(\frac{10}{3})$.
Now, how big is $\log_3(\frac{10}{3})$? Since $\frac{10}{3}$ is approximately 3.33, and $\log_3 3 = 1$, the exponent for c must be slightly greater than 1.

Let's summarize and compare:
* Exponent of a: $\log_2 3.4$. We know this is > 1. * Exponent of b: $\log_4 3.6$. We know this is < 1. * Exponent of c: $\log_3(\frac{10}{3})$. We know this is > 1.
From this, I can immediately tell that b is the smallest, since its exponent is the only one less than 1. Now the real question is: which is bigger, a or c? This means I need to compare their exponents: $\log_2 3.4$ and $\log_3(\frac{10}{3})$.
This is tricky. They have different bases. How can I compare them without a calculator? Maybe I can compare them to a common value, like 1.5?
Let's check the exponent of a: $\log_2 3.4$. Is it greater or less than 1.5? $1.5 = \frac{3}{2}$. So, I'm comparing $\log_2 3.4$ with $\frac{3}{2}$. This is the same as comparing 3.4 with $2^{3/2}$.
What is $2^{3/2}$? It's $(\sqrt{2})^3 \approx (1.414)^3$, which is hard to calculate. Let's try another way. Let's compare 3.4^2 with $(2^{3/2})^2 = 2^3 = 8$.
$3.4^2 = 11.56$. Since $11.56 > 8$, it means $3.4 > 2^{3/2}$. Therefore, $\log_2 3.4 > \frac{3}{2} = 1.5$.
Now let's check the exponent of c: $\log_3(\frac{10}{3})$. Is it greater or less than 1.5?
I'm comparing $\log_3(\frac{10}{3})$ with $\frac{3}{2}$. This is the same as comparing $\frac{10}{3}$ with $3^{3/2}$.
Let's square both sides again. I'm comparing $(\frac{10}{3})^2$ with $(3^{3/2})^2 = 3^3 = 27$.
$(\frac{10}{3})^2 = \frac{100}{9} \approx 11.11$.
Since $11.11 < 27$, it means $\frac{10}{3} < 3^{3/2}$. Therefore, $\log_3(\frac{10}{3}) < \frac{3}{2} = 1.5$.
This is fantastic! * Exponent of a is > 1.5.
* Exponent of c is < 1.5 (but we know it's > 1).
* Exponent of b is < 1.

So, the order of the exponents is: $(\log_2 3.4) > (\log_3(\frac{10}{3})) > (\log_4 3.6)$.
Since the base is 5 for all of them, the order of the numbers a, c, b must be the same as the order of their exponents.
Therefore, $a > c > b$.

Table 15: Cases generated by Qwen2.5-7B-Instruct in Math.