Exploration-Exploitation Prompting: A Dual-Process Framework for Complex Mathematical Problem Solving

Aniket Deroy[1]

***Abstract*— Solving high-dimensional mathematical problems requires more than just sequential reasoning; it requires a strategic balance between breadth of search and depth of computation. This paper introduces the Exploration-Exploitation Prompting (EEP) strategy. Inspired by reinforcement learning's Multi-Armed Bandit problem and human cognitive dual-process theory, EEP bifurcates the Large Language Model's (LLM) reasoning into two distinct phases: a "Global Exploration" phase to map potential solution spaces and a "Local Exploitation" phase to refine and execute the most promising path. We observe that our proposed Exploration-exploitation prompting method provides better results than the chain-of-thoughts and tree-of-thoughts prompting method on the MathNET dataset.**

I. Introduction

Current prompting techniques such as *Chain-of-Thought* (CoT) [1] and *Tree-of-Thoughts* (ToT) [2] often suffer from "greedy decoding" errors. Once a model commits to an initial incorrect mathematical assumption, it tends to hallucinate logic to support that path to maintain internal consistency [3]. EEP addresses this by forcing the model to act as both a **Scout** (Explorer) and a **Specialist** (Exploiter), decoupling strategy formulation from arithmetic execution [4]. The pursuit of human-level reasoning in artificial intelligence has found its most rigorous testing ground in mathematics. While LLMs have achieved remarkable success in natural language tasks, the "brittleness" of their reasoning becomes apparent in multi-step mathematical proofs. The prevailing standard, Chain-of-Thought (CoT), relies on a linear progression where the model predicts the next token based on the preceding context [5]. To address these structural limitations, researchers introduced Tree-of-Thoughts (ToT), which allows for broader search spaces [6]. However, ToT often lacks the specific "exploitation" depth required for the dense, symbolic environments of undergraduate and competition-level mathematics [7]. This paper proposes the Exploration-Exploitation Prompting (EEP) framework as a more robust alternative. EEP is rooted in the Dual-Process Theory of human cognition, distinguishing between divergent search (System 2) and convergent execution (System 1).By implementing a formal "Exploration" [9] phase, the model is prohibited from beginning its derivation until it has explored a diversified strategy space. We observe that our proposed Exploration-exploitation [10] prompting method provides better results than the chain-of-thoughts and tree-of-thoughts pompting method on the MathNET dataset [8].

*This work was not supported by any organization

[1]Electrical Engineering Department, IIT Delhi roydanik18kgpian.iitkgp.ac.in

II. Dataset

Modern high-level mathematical datasets represent a significant leap from basic arithmetic benchmarks, focusing instead on the structural complexity of formal proofs, symbolic logic, and competition-level problem-solving. At the forefront of this evolution is the MathNet [8] benchmark, which has recently set a new standard by aggregating over 30,000 problems from elite global competitions across 143 distinct categories. Unlike traditional datasets, MathNet evaluates a model's ability to recognize abstract structural similarities between seemingly unrelated problems, challenging systems to move beyond pattern matching toward genuine mathematical intuition. We randomly select 20% of mathematics problem from the MATHNET dataset [8].

III. Methodology

The EEP strategy is grounded in the logic of the **Upper Confidence Bound (UCB)**, adapted for linguistic reasoning and symbolic manipulation.

The two-phase strategy begins with Global Exploration, a stage where the model generates multiple diverse heuristic approaches—such as identifying symmetries, transformations, or geometric interpretations—to prevent premature "lock-in" to a single, potentially incorrect formula. By listing distinct mathematical frameworks before performing any calculations, the model maintains a high-level perspective on the problem. This is followed by Local Exploitation, where the model evaluates the utility of these approaches, selects the most viable path, and executes a rigorous, step-by-step derivation. This second phase prioritizes precision and exhaustive computation, specifically focusing on verifying dimensional consistency and checking for errors to ensure a successful final result. Table I represents the exploration-exploitation prompting technique for solving mathematical problems. Table III represents the exploration-exploitation prompts for mathematical problem solving via Exploration-exploitation prompting technique.

TABLE I

EXPLORATION-EXPLOITATION PROMPTING (EEP) FRAMEWORK

Phase	Prompt Component / Instruction
Exploration	Propose 3 diverse mathematical strategies (e.g., Algebraic, Geometric, or Combinatorial) for the problem: [INSERT PROBLEM]. Identify the logical bridge for each.
Evaluation	Assign a utility score (1-10) to each strategy based on its elegance and likelihood of success. Identify potential logical pitfalls.
Exploitation	Select the highest-scoring strategy and provide a rigorous, step-by-step L^AT_EX derivation. Revert to secondary strategies if a contradiction occurs.
Verification	Verify the result against boundary conditions and perform a consistency check on all intermediate lemmas.

Algorithm 1 Exploration-Exploitation Prompting (EEP)

Require: Complex Mathematical Problem P
Ensure: Formal Proof and Verified Solution Ψ
Phase I: Divergent Exploration (The Scout)
Generate a set of candidate strategies $\mathcal{S} \leftarrow \{s_1, s_2, \ldots, s_n\}$
for each $s_i \in \mathcal{S}$ **do**
  Define logical trajectory T_i and identify potential bottlenecks.
end for
Phase II: Utility Assessment (The Critic)
Assign weights w_i based on $w_i = f(Simplicity, Consistency)$
Select optimal strategy $s^* = argmax_{s_i \in \mathcal{S}}(w_i)$
Phase III: Convergent Exploitation (The Specialist)
while Solution not found **do**
  Execute s^* using step-by-step LaTeX derivation.
  if Logical contradiction $\perp$ is detected **then**
    Backtrack to $\mathcal{S}$ and select next s_j where $w_j = \max(w \setminus w_i)$
  else
    $\Psi \leftarrow CurrentDerivation$
    break
  end if
end while
Phase IV: Global Verification (The Auditor)
Cross-check Ψ against boundary conditions and n-case limits.
return Ψ =0

Model	Solution	Reasoning	Correlation
DeepSeek-R1			
DeepSeek-COT	0.205	0.231	0.455
DeepSeek-ToT	0.229	0.232	0.477
DeepSeek-EE(proposed)	**0.246**	**0.255**	0.557
Llama-3.2			
Llama-COT	0.215	0.239	0.455
Llama-ToT	0.236	0.239	0.488
Llama-EE(proposed)	**0.266**	**0.276**	**0.614**
Qwen-32B			
Qwen-COT	0.234	0.280	0.489
Qwen-ToT	0.259	0.277	0.503
Qwen-EE(proposed)	**0.277**	**0.298**	**0.665**
MathCoder-7B			
mathcoder-COT	0.218	0.256	0.459
mathcoder-ToT	0.255	0.265	0.498
mathcoder-EE(proposed)	**0.269**	**0.286**	**0.655**

TABLE II

STANDARD SOLUTION, REASONING, CORRELATION METRICS ON MATHNET DATASET OVER THREE LLMS-DEESPSEEK-R1, LLAMA-3.2, MATHCODER-7B,QWEN-32B

IV. RESULTS

The table II demonstrates that the EE (Exploration-Exploitation) framework consistently outperforms traditional Chain-of-Thought (CoT) and Tree-of-Thought (ToT) methods across all tested models. While ToT offers a marginal improvement over the CoT baseline, the EE framework provides a significant leap in performance, particularly in the Correlation metric, which measures how well a model's internal reasoning aligns with its final answer.

Among the specific models, MathCoder-7B shows the most substantial improvement, with its correlation score jumping from 0.459 to 0.655 when using the proposed method. Similarly, Llama-3.2 achieves its highest reasoning quality (0.276) and correlation (0.614) under the EE framework. Overall, the results indicate that the EE approach effectively bridges the gap between accurate results and sound logical steps, leading to more reliable and mathematically consistent outputs across diverse LLM architectures.The proposed Qwen-EE framework consistently outperforms both COT and ToT across all metrics, achieving its most significant breakthrough with a 32.2

V. CONCLUSION

This paper introduced Exploration-Exploitation Prompting (EEP), a novel framework designed to mitigate the inherent "greedy decoding" limitations of current Large Language Models in complex mathematical reasoning. By decoupling the strategic mapping of solution spaces from formal execution, EEP effectively simulates human-like cognitive dual-processing. Our experimental results across diverse architectures—DeepSeek-R1, Llama-3.2, and MathCoder-7B, Qwen-32B — consistently demonstrate that the EEP framework outperforms both Chain-of-Thought (CoT) and Tree-of-Thought (ToT) methodologies.

REFERENCES

[1] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837

[2] Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. Advances in neural information processing systems, 36, 11809-11822.

[3] Schweinsberg, R. Z. (2020). Examining the Internal Consistency and Predictive Validity of a Post-Test Administered to Developmental Math Students.

[4] Cartwright, R., Boehm, B. (1990). Exact Real Arithmetic, formulating real numbers as functions. Research Topics in Functional Programming. University of Texas at Austin Year of Programming Series, 43-64.

[5] Pal, K., Sun, J., Yuan, A., Wallace, B. C., Bau, D. (2023, December). Future lens: Anticipating subsequent tokens from a single hidden state. In Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL) (pp. 548-560).

[6] Zhang, W. (1999). State-space search: Algorithms, complexity, extensions, and applications. Springer Science Business Media.

[7] Xu, X., Zhang, J., Chen, T., Chao, Z., Hu, J., Yang, C. (2025). Ugmathbench: A diverse and dynamic benchmark for undergraduate-level mathematical reasoning with large language models. arXiv preprint arXiv:2501.13766.

[8] Alshammari, S., Wen, K., Zainal, A., Hamilton, M., Safaei, N., Albarakati, S., ... Torralba, A. (2026). MathNet: a Global Multimodal Benchmark for Mathematical Reasoning and Retrieval. arXiv preprint arXiv:2604.18584.

[9] Bruhn, D., Manzella, A., Vuataz, F., Faulds, J., Moeck, I., Erbas, K. (2010). Exploration methods. Geothermal energy systems: exploration, development, and utilization, 37-112.

[10] Cheng, L., Ahmed, S., Liljestrand, H., Nyman, T., Cai, H., Jaeger, T., ... Yao, D. (2021). Exploitation techniques for data-oriented attacks with existing and potential defense approaches. ACM Transactions on Privacy and Security (TOPS), 24(4), 1-36.