A Hippocampal-PFC Inspired Neuro-Symbolic Architecture for Contextually Anchored Question Chain Generation

Anonymous ACL submission

Abstract

Generating high-quality, interconnected questions remains a significant challenge in artificial intelligence (AI), particularly in applications requiring logical coherence and social relevance. Current methods often lack a cognitive foundation to ensure meaningful question relationships, limiting their effectiveness in dynamic environments. To address this gap, we propose a novel neuroscience-inspired framework, HPN-SCA, that integrates AI with theories of prefrontal cortex function, hippocampal memory retrieval, and the dynamic interplay between the Default Mode Network (DMN) and Central Executive Network (CEN). Our methodology consists of three key steps: (1) the Prefrontal Cortex Simulator, where dual models emulate the dorsolateral prefrontal cortex (DLPFC) for logical structuring and the ventromedial prefrontal cortex (VMPFC) for social contextualization to generate preliminary questions; (2) the Hippocampus Simulator, which classifies questions into Scenario-based (retrieved from knowledge bases) or Logic-based (interactively generated) chunks, mimicking memory association mechanisms; and (3) the DMN/CEN Simulator, where difficulty-based routing refines questions through either associative (DMN) or rigorous (CEN) processing. Experiments show our HPN-SCA method outperforms baselines in coherence, diversity, and human evaluation. This work integrates AI and cognitive science, enabling applications in education and conversational AI. Future work will explore additional cognitive mechanisms.

1 Introduction

Automatic question generation remains a fundamental challenge in artificial intelligence, with significant implications for educational technologies, conversational systems, and knowledge discovery. While recent advances in natural language processing have improved question generation capabilities, current approaches often produce isolated questions lacking logical coherence, contextual depth, or social relevance. This limitation stems from their failure to incorporate the cognitive mechanisms underlying human question formulation.

Existing methods predominantly rely on pattern recognition from large text corpora (Liu et al., 2025b; Xie et al., 2025) or sequence-to-sequence architectures (Sujatha et al., 2025; Bi et al., 2025), overlooking the neurocognitive processes that enable humans to generate meaningful, interconnected questions. Particularly absent are models that account for: (1) the complementary roles of dorsolateral and ventromedial prefrontal cortices in logical structuring and social contextualization (Guo and Shing, 2025; Badre and Nee, 2018), (2) hippocampal mechanisms for memory retrieval and association (Hasselmo and Eichenbaum, 2005), and (3) the dynamic interaction between Default Mode and Central Executive Networks during question refinement (Chen et al., 2013). This cognitive gap limits both the quality and applicability of generated questions.

2 Related Work

Question Generation Methods in AI: Existing approaches to automated question generation can be categorized into three main paradigms. Rule-based systems (Caufield et al., 2024; Lyu et al., 2025) employ syntactic patterns and template matching to transform declarative sentences into questions, offering high precision but limited scalability. Statistical methods (Rani and Jain, 2024; Lin, 2024) utilize n-gram models and semantic role labeling, improving flexibility but struggling with complex sentence structures. Recent neural approaches (Raiaan et al., 2024; Annepaka and Pakray, 2024; Veeramachaneni, 2025) leverage sequence-to-sequence architectures with attention mechanisms, achieving state-of-the-art performance through transformer-based models like

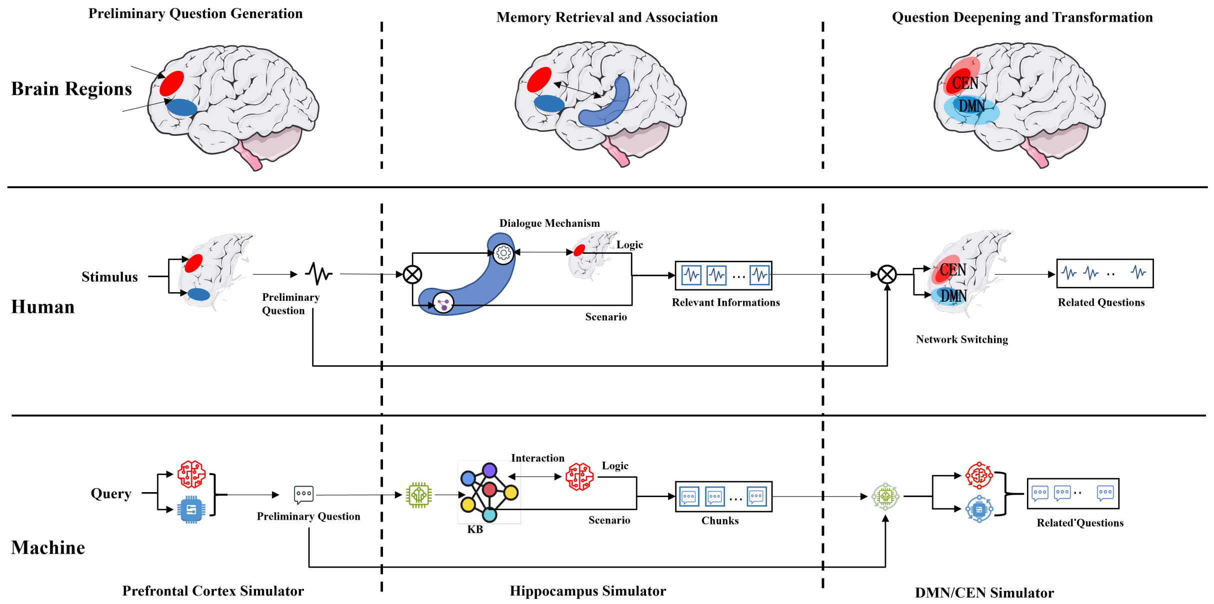

Figure 1: Overview of the HPN-SCA Framework: The HPN-SCA framework integrates AI with neuroscience to generate high - quality questions. It has three steps: the Prefrontal Cortex Simulator creates Preliminary Questions by mimicking prefrontal cortex functions; the Hippocampus Simulator classifies and retrieves/generates Chunks like the hippocampus's memory processes; and the DMN/CEN Simulator refines questions by routing Chunks to different models based on difficulty, simulating network switching in the brain.

BERT and GPT. While these neural methods generate fluent questions, they often produce isolated queries without considering logical sequences or contextual relationships (Zhou et al., 2024).

Neuroscience-Inspired AI Research: Several studies have successfully integrated cognitive principles into AI systems. Memory-augmented networks (Jimenez Gutierrez et al., 2024; Zhu et al., 2025) have modeled hippocampal functions for question answering, while prefrontal cortex simulations (Wei et al., 2025) have enhanced decision-making systems. Particularly relevant are: (1) DLPFC-inspired architectures for task organization (Kahnt et al., 2011), (2) VMPFC models for social cognition (Labutina et al., 2024), and (3) DMN/CEN interaction systems for creative problem solving (Maslova et al., 2025). However, these implementations remain specialized - no existing work combines these cognitive components for question generation. Recent hybrid systems (Liu et al., 2025a; Zheng et al., 2025) have shown promise in integrating multiple brain regions, but focus primarily on memory retrieval rather than generative tasks.

Research Gaps and Our Position: Three critical gaps emerge from this review: First, current question generation systems lack mechanisms for maintaining logical coherence across multiple questions, analogous to the DLPFC's organizational function. Second, they fail to incorporate social-relevance filtering comparable to VMPFC operations. Third, no existing framework implements the dynamic DMN/CEN switching crucial for adapting question difficulty and type. Our work addresses these limitations through: (1) simultaneous DLPFC/VMPFC simulation for balanced question formulation, (2) hippocampal modeling for interconnected question chunks, and (3) difficulty-adaptive DMN/CEN routing - representing the first unified cognitive architecture for comprehensive question generation.

The main contributions of our paper are as follows:

- Methodological Innovation: Emphasize the novelty of combining artificial intelligence techniques with neuroscience concepts to model different brain regions (prefrontal cortex, hippocampus, DMN/CEN) for question - generation in the HPN-SCA framework. Explain how this integration provides a new perspective on question - generation.

- Performance Improvement: Present how the HPN-SCA framework outperforms existing methods in terms of question quality and interconnectedness. Provide quantitative and qualitative metrics to support this claim.

- Theoretical and Practical Implications: Discuss the theoretical contribution to the understanding of question - generation and its underlying mechanisms within the HPN-SCA framework, as well as the practical implications for applications in different domains.

3 Methods

3.1 Overall Framework

The HPN-SCA Framework comprises three interconnected modules (Figure 1): (1) Prefrontal Cortex Simulator for initial question formulation, (2) Hippocampus Simulator for memory-based question association, and (3) DMN/CEN Simulator for adaptive question refinement. The system takes an input query Q and progressively transforms it through these modules to output a set of interconnected questions $\{q_1, ..., q_n\}$ with balanced logical structure and social relevance.

3.2 Probabilistic Formalization of the HPN-SCA Framework

We model the question-generation process as a **probabilistic pipeline** where each component (Prefrontal Cortex, Hippocampus, and DMN/CEN Simulators) contributes to the generation of high-quality, interconnected questions. Below, we formalize each step with probabilistic formulations and define the associated parameters.

3.2.1 Prefrontal Cortex Simulator

The **DLPFC** and **VMPFC** simulators generate a preliminary question Q_{prelim} from an input query q.

Formulation The combined generation process is modeled as:

$$P(Q_{\text{prelim}} \mid q) = \alpha \cdot P_{\text{DLPFC}}(Q_{\text{prelim}} \mid q) + (1 - \alpha) \cdot P_{\text{VMPFC}}(Q_{\text{prelim}} \mid q) \quad (1)$$

where:

- $P_{\text{DLPFC}}(Q_{\text{prelim}} \mid q)$: Probability of generating Q_{prelim} under logical structuring (DLPFC).
- $P_{\text{VMPFC}}(Q_{\text{prelim}} \mid q)$: Probability of generating Q_{prelim} under social-contextual refinement (VMPFC).
- $\alpha \in [0, 1]$: Weight balancing logical vs. social considerations (tuned empirically).

Explanation:

- The DLPFC model generates questions with high P_{DLPFC} when logical coherence is prioritized.
- The VMPFC model increases P_{VMPFC} when social relevance is needed.
- α controls the trade-off (e.g., $\alpha = 0.7$ for exam-style questions, $\alpha = 0.3$ for conversational questions).

3.3 Hippocampus Simulator

The **Hippocampus Simulator** classifies Q_{prelim} into **Scenario (S)** or **Logic (L)** categories and retrieves/generates associated knowledge chunks C.

The classification and chunk generation follow:

$$P(S \mid Q_{\text{prelim}}) = \sigma(f_{\text{class}}(Q_{\text{prelim}})) \quad (2)$$

$$P(L \mid Q_{\text{prelim}}) = 1 - P(S \mid Q_{\text{prelim}}) \quad (3)$$

where:

- $\sigma(\cdot)$: Sigmoid function for binary classification.
- f_{class}: A language model-based classifier (e.g., few-shot LLM scoring).

For chunk generation:

- **If Q_{prelim} is Scenario-type:**

$$C \sim P_{\text{retrieve}}(C \mid Q_{\text{prelim}}, \mathcal{K}) \quad (4)$$

where $\mathcal{K}$ is the knowledge base, and retrieval follows a similarity-based distribution (e.g., $P_{\text{retrieve}} \propto \text{sim}(Q_{\text{prelim}}, C)$).

- **If Q_{prelim} is Logic-type:**

$$C \sim P_{\text{DLPFC}}(C \mid Q_{\text{prelim}}) \quad (5)$$

where the DLPFC simulator generates chunks via few-shot prompting.

Explanation:

- The classifier f_{class} determines whether the question requires factual recall (Scenario) or logical derivation (Logic).
- P_{retrieve} uses embeddings (e.g., cosine similarity) to fetch relevant chunks.
- P_{DLPFC} ensures logical continuity by reusing the DLPFC's structured generation.

3.4 DMN/CEN Simulator

The **DMN (Default Mode Network)** and **CEN (Central Executive Network)** simulators refine chunks C into related questions Q_{related} based on difficulty.

First, difficulty is classified:

$$P(\text{Hard} \mid C) = \sigma(g_{\text{diff}}(C)) \tag{6}$$

where g_{diff} is a language model-based difficulty scorer.

Then, Q_{related} is generated via:

$$\begin{aligned} P(Q_{\text{related}} \mid C) &= \beta \cdot P_{\text{CEN}}(Q_{\text{related}} \mid C) \\ &\quad + (1-\beta) \cdot P_{\text{DMN}}(Q_{\text{related}} \mid C) \end{aligned} \tag{7}$$

where:

- $P_{\text{CEN}}(Q_{\text{related}} \mid C)$: Analytical refinement (high-depth reasoning).
- $P_{\text{DMN}}(Q_{\text{related}} \mid C)$: Associative/creative expansion (broad connections).
- $\beta = P(\text{Hard} \mid C)$: Dynamic weight favoring CEN for hard chunks.

Explanation:

- g_{diff} assigns difficulty (e.g., based on chunk complexity or ambiguity).
- Hard chunks ($\beta \approx 1$) use **CEN** for rigorous question decomposition.
- Easy chunks ($\beta \approx 0$) use **DMN** for creative associations (e.g., analogies).

3.5 Full Pipeline Integration

The end-to-end generation of related questions follows:

$$\begin{aligned} P(Q_{\text{related}} \mid q) = \sum_{Q_{\text{prelim}}, C} & P(Q_{\text{related}} \mid C) \\ & \cdot P(C \mid Q_{\text{prelim}}) \cdot P(Q_{\text{prelim}} \mid q) \end{aligned} \tag{8}$$

The pseudo code of the HPN-SCA Framework algorithm is shown in Algorithm 1.

4 Experiments

4.1 Experimental Setup

In the experiment setups of our study, the selection of models is crucial for achieving accurate and effective results. For the main model, we have

Algorithm 1 Question Generation Algorithm of HPN-SCA Framework

Require: Input query q
Ensure: Set of interconnected questions $\{q_1, \ldots, q_n\}$
// Prefrontal Cortex Simulator
Define $\alpha \in [0, 1]$ (weight for logical vs. social considerations)
$P_{\text{DLPFC}}(Q_{\text{prelim}} \mid q)$:= Probability of generating Q_{prelim} under logical structuring (using DLPFC model)
$P_{\text{VMPFC}}(Q_{\text{prelim}} \mid q)$:= Probability of generating Q_{prelim} under social - contextual refinement (using VMPFC model)
$Q_{\text{prelim}} := \alpha \cdot P_{\text{DLPFC}}(Q_{\text{prelim}} \mid q) + (1-\alpha) \cdot P_{\text{VMPFC}}(Q_{\text{prelim}} \mid q)$
// Hippocampus Simulator
f_{class} := Language model - based classifier
$P(S \mid Q_{\text{prelim}}) := \sigma(f_{\text{class}}(Q_{\text{prelim}}))$ (probability of Q_{prelim} being Scenario - type)
$P(L \mid Q_{\text{prelim}}) := 1 - P(S \mid Q_{\text{prelim}})$ (probability of Q_{prelim} being Logic - type)
if $P(S \mid Q_{\text{prelim}})$ is high **then**
    $\mathcal{K}$:= Knowledge base
    $C \sim P_{\text{retrieve}}(C \mid Q_{\text{prelim}}, \mathcal{K})$ (retrieve chunks based on similarity to Q_{prelim} from $\mathcal{K}$)
else
    $C \sim P_{\text{DLPFC}}(C \mid Q_{\text{prelim}})$ (generate chunks using DLPFC model)
end if
// DMN/CEN Simulator
g_{diff} := Language model - based difficulty scorer
$P(\text{Hard} \mid C) := \sigma(g_{\text{diff}}(C))$ (probability of chunk C being hard)
$\beta := P(\text{Hard} \mid C)$
$P_{\text{CEN}}(Q_{\text{related}} \mid C)$:= Probability of generating Q_{related} via analytical refinement (using CEN model)
$P_{\text{DMN}}(Q_{\text{related}} \mid C)$:= Probability of generating Q_{related} via associative/creative expansion (using DMN model)
$Q_{\text{related}} := \beta \cdot P_{\text{CEN}}(Q_{\text{related}} \mid C) + (1-\beta) \cdot P_{\text{DMN}}(Q_{\text{related}} \mid C)$
Return Q_{related} as the set of interconnected questions $\{q_1, \ldots, q_n\}$

Table 1: Description of Artificial Intelligence-Related Datasets

Dataset Name	Type	Key Features	Scale
CoQA (Reddy et al., 2019)	Conversational Question Answering	Conversational questions; free-form text answers; evidence subsequences highlighted; 7 diverse domains; includes coreference and pragmatic reasoning challenges	127,000 questions from 8,000 conversations
MCTest (Richardson et al., 2013)	Machine Comprehension	Freely available; gathered via Mechanical Turk	660 stories with associated questions
High School Chinese Educational (HSCE)[1]	Educational Questions	20 different texts with learning tasks and questions	484 learning tasks, 1,452 questions

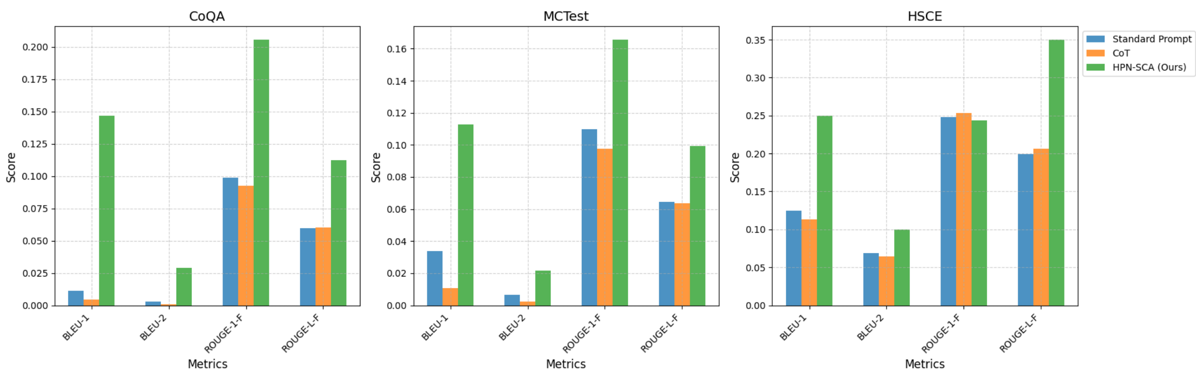

Figure 2: Performance Comparison on Three Datasets

chosen Qwen - 14B - chat(Bai et al., 2023). This model was selected due to its distinct advantages in handling cross - text datasets. Its architecture and training process endow it with the ability to understand and process text across diverse sources, which is essential for the tasks at hand. In addition, for the similarity comparison in the retrieval process, we employed the Sbert model(Reimers and Gurevych, 2019). Sbert is well - known for its efficiency and accuracy in calculating semantic similarities between text snippets. By using this model, we can effectively retrieve relevant information from a large corpus, enabling us to make more informed comparisons and analyses within our experimental framework.

4.1.1 Datasets

We evaluate our the HPN-SCA Framework on three benchmark datasets. The details are presented in Table 1.

[1]Data source: https://anonymous.4open.science/r/HSCE-ED7B

4.2 Quantitative Results

In this section, we present the performance comparison of different methods on three datasets: CoQA, MCTest, and HSCE, using evaluation metrics such as BLEU-1, BLEU-2, ROUGE-1-F, and ROUGE-L-F, with the results summarized in Figure 2. This study assesses the question generation performance of the HPN-SCA method, comparing it against Standard Prompt and CoT(Wei et al., 2022) methods. Across all datasets, HPN-SCA outperforms the others. On CoQA and MCTest, it demonstrates significant improvements in all metrics, achieving 12 - 15 times higher scores than the baseline in some cases. For HSCE, although CoT has a relatively high ROUGE-1-F score, HPN-SCA dominates in BLEU metrics and attains the highest ROUGE-L-F score. The core advantages of HPN-SCA, including its multi-dimensional semantic modeling, robustness across datasets, and superiority in handling complex contexts, allow it to capture semantic relationships, maintain stable performance, and generate contextually consistent questions, thereby providing a new paradigm for natural language tasks.

4.3 Hyperparameter Influence

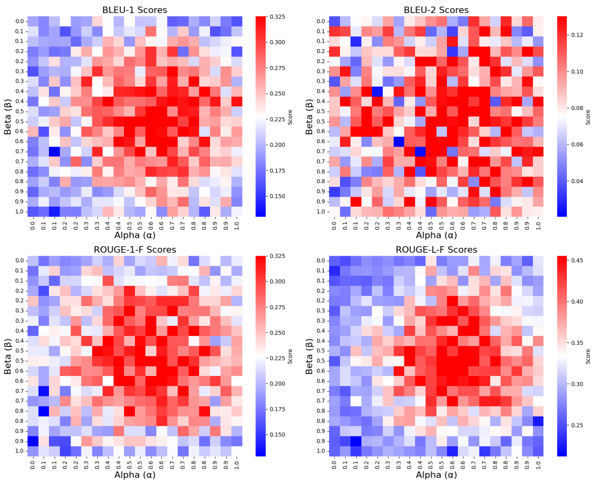

Figure 3: Hyperparameter vs. Evaluation Metrics Heatmap

The heatmap (Figure 3) analysis reveals that optimal performance across all metrics (BLEU-1, BLEU-2, ROUGE-1-F, and ROUGE-L-F) occurs at intermediate values of α (0.5–0.7) and β (0.4–0.6), demonstrating the importance of balanced trade-offs between logical structuring (α) and difficulty-based refinement (β). Extreme values in either hyperparameter degrade performance, with α showing stronger influence than β, indicating that neither purely logical (DLPFC-dominated) nor purely social-contextual (VMPFC-dominated) generation is optimal. ROUGE-L-F exhibits the highest sensitivity to hyperparameter tuning, suggesting its stronger dependence on the framework's balanced refinement strategy compared to lexical overlap metrics (BLEU-1/2). These results empirically validate the probabilistic formulation of HPN-SCA, highlighting that joint optimization of α and β is crucial for maximizing question-generation quality across diverse evaluation dimensions.

4.4 Qualitative Analysis

4.4.1 Case Study

Input: *"Learning Task: Writing a Biography The first section of the exhibit board requires an introduction to Tu Youyou. Please draft a biographical sketch to help students quickly understand her significance. Students should carefully read the assigned text, gather and review relevant materials about Tu Youyou, and present the information objectively and rigorously, with a minimum length of 200 words."*

Output Questions:

1. (Standard Prompt) *Based on the content of the text Artemisinin, compose a biographical introduction for Tu Youyou to be displayed on the exhibit board. Your writing should draw upon the text's descriptions of her research background, key contributions, career trajectory, and personal qualities, ensuring factual accuracy while maintaining an objective and scholarly tone. The summary should be concise yet comprehensive, with a minimum of 200 words. After completing the reading and organizing your materials, draft a well-rounded overview.*

2. (CoT) *Who is Tu Youyou, and what is her significance in the history of Chinese medicine? In which field did Tu Youyou make her most notable contributions? In what year was she born, and which province did she come from? What was her pivotal breakthrough in scientific exploration? For which discovery or invention is Tu Youyou most widely recognized? What challenges and obstacles did she face during her research? How has her work impacted global healthcare? What major scientific awards or honors has Tu Youyou received? Which traditional herbal remedy inspired her research team's extraction process for artemisinin? How do Tu Youyou's personal traits—such as perseverance and innovative thinking—manifest in her scientific journey? What lessons can contemporary scholars and scientists draw from Tu Youyou's life and work? How does Tu Youyou's story inspire young people to pursue scientific aspirations? These questions are designed to guide students in closely analyzing the text, seeking answers, and composing the biography.*

3. (HPN-SCA) *Before reading the text, list Tu Youyou's key contributions to the discovery of artemisinin. Which paragraph(s) detail Tu Youyou's research background? What professional competencies does this highlight? What challenges did Tu Youyou encounter during her research, and how did she address them? What was the "523 Project," and how did it influence Tu Youyou's work? Identify*

Table 2: Manual Scoring Rubric

Criterion	Score Range	Evaluation Description
Expression Quality	8-10	Flawless grammar, coherent semantics, scientifically accurate
	5-7	Minor grammatical issues, basically clear, mostly accurate
	1-4	Confused structure, ambiguous meaning, cognitive errors
Content Quality	8-10	Highly scenario-adaptive, strong emotional resonance
	5-7	Partially scenario-adaptive, moderate practical value
	1-4	Poor scenario fit, low practical value, potential discomfort
Question Relevanc	8-10	Strict thematic focus, progressive logic, highly complementary
	5-7	Basic thematic consistency, some logical repetition
	1-4	Dispersed themes, chaotic logic, redundant content

Table 3: Human Evaluation

Data Set	Method	Expression Quality	Content Quality	Question Relevance
CoQA	Standard Prompt	7.47	4.20	3.87
	CoT	7.53	4.267	4.13
	HPN-SCA	**8.07**	**7.93**	**8.20**
MCTest	Standard Prompt	7.27	4.40	3.87
	CoT	7.20	4.47	4.33
	HPN-SCA	**7.93**	**8.13**	**8.07**
HSCE	Standard Prompt	4.71	4.19	2.42
	CoT	4.33	3.00	2.33
	HPN-SCA	**7.10**	**7.14**	**7.00**

the traditional herbal extraction method mentioned in the text and explain its underlying principles. How many critical turning points occurred during the artemisinin extraction process? Describe these pivotal events. Interpret Tu Youyou's statement, "Inspiration came from The Handbook of Prescriptions for Emergencies." Does this reflect her innovative integration of traditional knowledge? What criticisms or difficulties did Tu Youyou face upon receiving awards? How did her resilience shine through? Which of Tu Youyou's qualities or values are most worthy of emulation? Provide textual examples. When drafting the biography, which key achievements and traits of Tu Youyou would you emphasize, based on the text? These questions aim to guide students in examining Tu Youyou's research process and personal attributes from multiple perspectives, ensuring thorough preparation for writing the biography.

4.5 Human Evaluation

To systematically evaluate the quality of questions across multiple dimensions, we propose a manual scoring rubric with three key criteria: Expression Quality, Content Quality, and Question Relevance. Each criterion is assessed on a 10-point scale, with higher scores indicating stronger adherence to linguistic precision, practical utility, and logical coherence. The detailed scoring guidelines are presented in Table 2, which provides clear benchmarks for distinguishing between excellent (8–10), satisfactory (5–7), and inadequate (1–4) performance. This rubric ensures consistent and transparent evaluation while accommodating nuanced variations in question design. The human evaluation results are presented in Table 3.

The results show that across different datasets (CoQA, MCTest, HSCE), the HPN - SCA method outperforms the Standard Prompt and CoT methods in all three evaluation criteria: Expression Quality, Content Quality, and Question Relevance. Specifically, on CoQA and MCTest, HPN - SCA achieves scores above 7.9 in all criteria, demonstrating excellent performance in grammar, scenario adaptability, and thematic focus. For the HSCE dataset, although the base scores of the other two methods are lower, HPN - SCA still significantly improves each indicator to around 7. This indicates that HPN - SCA has strong advantages in the question generation task, especially in enhancing the quality and relevance of generated questions.

Limitations

While HPN-SCA demonstrates promising results in generating contextually anchored question chains, several limitations warrant discussion. First, the current architecture relies on predefined knowledge bases for scenario-based question retrieval, which may limit generalization in domains with sparse or evolving knowledge. Second, the cognitive simulations (DLPFC/VMPFC, DMN/CEN dynamics) are abstract computational approximations rather than biologically precise models, potentially overlooking nuanced neural mechanisms. Third, the evaluation metrics, though comprehensive, do not fully capture longitudinal learning effects that hippocampal-prefrontal interactions would enable in biological systems. Addressing these limitations through adaptive knowledge graphs, more detailed neural circuit modeling, and optimized inference pipelines represents key directions for future research.

Acknowledgments

None

References

Yadagiri Annepaka and Partha Pakray. 2024. Large language models: A survey of their development, capabilities, and applications. *Knowledge and Information Systems*, pages 1–56.

David Badre and Derek Evan Nee. 2018. Frontal cortex and the hierarchical control of behavior. *Trends in cognitive sciences*, 22(2):170–188.

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*.

Sheng Bi, Zeyi Miao, and Qizhi Min. 2025. Lemon: A knowledge-enhanced, type-constrained and grammar-guided model for question generation over knowledge graphs. *IEEE Transactions on Learning Technologies*.

J Harry Caufield, Harshad Hegde, Vincent Emonet, Nomi L Harris, Marcin P Joachimiak, Nicolas Matentzoglu, HyeongSik Kim, Sierra Moxon, Justin T Reese, Melissa A Haendel, and 1 others. 2024. Structured prompt interrogation and recursive extraction of semantics (spires): A method for populating knowledge bases using zero-shot learning. *Bioinformatics*, 40(3):btae104.

Ashley C Chen, Desmond J Oathes, Catie Chang, Travis Bradley, Zheng-Wei Zhou, Leanne M Williams, Gary H Glover, Karl Deisseroth, and Amit Etkin. 2013. Causal interactions between fronto-parietal central executive and default-mode networks in humans. *Proceedings of the National Academy of Sciences*, 110(49):19944–19949.

Dingrong Guo and Yee Lee Shing. 2025. Linking the congruency effect in memory to confirmation bias in decision-making across the lifespan—common roles of the medial prefrontal cortex: A selective review. *European Psychologist*.

Michael E Hasselmo and Howard Eichenbaum. 2005. Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural networks*, 18(9):1172–1190.

Bernal Jimenez Gutierrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. 2024. Hipporag: Neurobiologically inspired long-term memory for large language models. *Advances in Neural Information Processing Systems*, 37:59532–59569.

Thorsten Kahnt, Jakob Heinzle, Soyoung Q. Park, and John-Dylan Haynes. 2011. Decoding different roles for vmpfc and dlpfc in multi-attribute decision making. *NeuroImage*, 56(2):709–715. Multivariate Decoding and Brain Reading.

Natalya Labutina, Sergey Polyakov, Liudmila Nemtyreva, Alina Shuldishova, and Olga Gizatullina. 2024. Neural correlates of social decision-making. *Iranian Journal of Psychiatry*, 19(1):148.

Sha Lin. 2024. Evaluating llms' grammatical error correction performance in learner chinese. *PloS one*, 19(10):e0312881.

Bang Liu, Xinfeng Li, Jiayi Zhang, Jinlin Wang, Tanjin He, Sirui Hong, Hongzhang Liu, Shaokun Zhang, Kaitao Song, Kunlun Zhu, and 1 others. 2025a. Advances and challenges in foundation agents: From brain-inspired intelligence to evolutionary, collaborative, and safe systems. *arXiv preprint arXiv:2504.01990*.

Jianyu Liu, Yi Huang, Sheng Bi, Junlan Feng, and Guilin Qi. 2025b. From superficial to deep: Integrating external knowledge for follow-up question generation using knowledge graph and LLM. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 828–840, Abu Dhabi, UAE. Association for Computational Linguistics.

Yuanjie Lyu, Zhiyu Li, Simin Niu, Feiyu Xiong, Bo Tang, Wenjin Wang, Hao Wu, Huanyong Liu, Tong Xu, and Enhong Chen. 2025. Crud-rag: A comprehensive chinese benchmark for retrieval-augmented generation of large language models. *ACM Transactions on Information Systems*, 43(2):1–32.

Olga Maslova, Natalia Shusharina, and Vasiliy Pyatin. 2025. The neurosociological paradigm of the metaverse. *Frontiers in Psychology*, 15:1371876.

Mohaimenul Azam Khan Raiaan, Md Saddam Hossain Mukta, Kaniz Fatema, Nur Mohammad Fahad, Sadman Sakib, Most Marufatul Jannat Mim, Jubaer Ahmad, Mohammed Eunus Ali, and Sami Azam. 2024. A review on large language models: Architectures, applications, taxonomies, open issues and challenges. *IEEE access*, 12:26839–26874.

Shweta Rani and Rhea Jain. 2024. Enhancing spoken text with punctuation prediction using n-gram language model in intelligent technical text processing software. In *Advancing Software Engineering Through AI, Federated Learning, and Large Language Models*, pages 201–217. IGI Global.

Siva Reddy, Danqi Chen, and Christopher D Manning. 2019. Coqa: A conversational question answering challenge. *Transactions of the Association for Computational Linguistics*, 7:249–266.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Matthew Richardson, Christopher JC Burges, and Erin Renshaw. 2013. Mctest: A challenge dataset for the open-domain machine comprehension of text. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 193–203.

B Sujatha, P Nagamani, N Leelavathy, and 1 others. 2025. Aslm-aqg: Attention-based sequence learning model for automatic question generation. In *Algorithms in Advanced Artificial Intelligence*, pages 459–464. CRC Press.

Vinod Veeramachaneni. 2025. Large language models: A comprehensive survey on architectures, applications, and challenges. *Advanced Innovations in Computer Programming Languages*, 7(1):20–39.

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.

Jinzhao Wei, Licong Li, Jiayi Zhang, Erdong Shi, Jianli Yang, and Xiuling Liu. 2025. Computational modeling of the prefrontal-cingulate cortex to investigate the role of coupling relationships for balancing emotion and cognition. *Neuroscience Bulletin*, 41(1):33–45.

Eric Xie, Guangzhi Xiong, Haolin Yang, Olivia Coleman, Michael Kennedy, and Aidong Zhang. 2025. Leveraging grounded large language models to automate educational presentation generation. In *Large Foundation Models for Educational Assessment*, pages 207–220. PMLR.

Qinghua Zheng, Huan Liu, Xiaoqing Zhang, Caixia Yan, Xiangyong Cao, Tieliang Gong, Yong-Jin Liu, Bin Shi, Zhen Peng, Xiaocen Fan, and 1 others. 2025. Machine memory intelligence: Inspired by human memory mechanisms. *Engineering*.

Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, and 1 others. 2024. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. *International Journal of Machine Learning and Cybernetics*, pages 1–65.

Yunlai Zhu, Yongjie Zhao, Junjie Zhang, Xi Sun, Ying Zhu, Xu Zhou, Xuming Shen, Zuyu Xu, Zuheng Wu, and Yuehua Dai. 2025. Memristor-based circuit design of interweaving mechanism of emotional memory in a hippocamp-brain emotion learning model. *Neural Networks*, 186:107276.

A Example Appendix

A.1 Prefrontal Cortex

A.1.1 DLPFC Prompt

```
"Act as logical question designer. Given $\{q\}$,
generate 3 variants that: 1) Maintain logical
coherence 2) Decompose complexity 3) Follow
deductive structures. Format: - Primary:
[Reformulation] - Follow-up 1/2: [Sub-questions]"
```

A.1.2 VMPFC Prompt

```
"Generate 3 social variants of $\{q\}$: 1) Cultural
references 2) Natural dialog structures 3)
Formality-adjusted versions. Output: - Professional:
[Formal] - Casual: [Conversational] - Adapted:
[Locale-specific]"
```

A.2 Hippocampus

A.2.1 Classifier Prompt

```
"Classify $\{Q_{\text{prelim}}\}$ as: Scenario (factual recall)
or Logic (derivation needed). Output: Type:
[S/L] Confidence: [0-100\%] Justification:
[1-sentence]"
```

A.2.2 Retrieval Prompt

```
"Generate knowledge chunks for $\{Q_{\text{prelim}}\}$.
Scenario-type: Factual excerpts. Logic-type:
1) Core Concept 2) Supporting Facts 3) Logical
Connections 4) Potential Gaps"
```

A.3 DMN/CEN

A.3.1 Difficulty Prompt

```
"Rate $\{C\}$ difficulty (1-5): 1) Conceptual
Complexity 2) Background Knowledge 3) Cognitive
Load. Thresholds: Hard (sum$\geq$10), Easy (sum$\leq$6)"
```

A.3.2 CEN Prompt

```
"Refine hard question $\{C\}$: 1) Identify assumptions
2) Decompose 3) Add constraints 4) Verification
methods"
```

A.3.3 DMN Prompt

```
"Generate creative associations for $\{C\}$: 1) Analogies
(3 domains) 2) Alternative phrasings 3) Cross-disciplinary
connections 4) Counterfactuals"
```