Creative Thought Embeddings: A Framework for Instilling Creativity in Large language Models

Qusay H. Mahmoud
Ontario Tech University
qusay.mahmoud@ontariotechu.ca

Abstract

Creative intelligence represents a critical frontier in artificial intelligence research. While modern large language models (LLMs) excel in logical reasoning and factual responses, they often produce outputs that are predictable and lack genuine originality. This paper introduces *Creative Thought Embeddings (CTE)*, a framework that embeds a creative bias directly into the latent representations of LLMs. By integrating a structured, multi-phase process that mirrors human divergent thinking, beginning with brainstorming and followed by synthesis, CTE guides models to generate outputs that are more novel, surprising, and contextually rich. The effectiveness of CTE is demonstrated across domains including humor generation, narrative storytelling, and educational explanations. Evaluation results, which employ quantitative lexical metrics and GPT-4o–based automated scoring show that while baseline models may exhibit greater surface-level lexical diversity, CTE enhances deeper semantic novelty and creative coherence. Finally, the paper presents a comparative analysis with standard prompt engineering and chain-of-thought approaches, discusses the trade-offs involved, and offers recommendations for further research and practical implementation.

Introduction

The remarkable progress of large language models (LLMs) in tasks that require logical reasoning and pattern recognition has raised expectations about their potential. However, when it comes to generating truly creative content, these models often fall short. Creativity is not merely a byproduct of fluency or grammatical correctness; it entails the ability to produce novel, useful, and surprising ideas, which is a capacity that remains elusive in current systems. Applications that demand creativity include storytelling, humor, design ideation, and educational innovation.

Standard prompting techniques and chain-of-thought (CoT) prompting guide LLMs through step-by-step reasoning, thereby improving performance on logical and arithmetic tasks. However, such methods tend to confine the models to a linear, predictable mode of operation that is not well suited for creative endeavors. Human creativity is characterized by imaginative leaps, the formation of unexpected associations, and the refinement of vague ideas into coherent insights. Accordingly, this paper introduces *Creative Thought Embeddings* (CTE), a framework that seeks to instill creativity in LLMs by embedding a creative bias within their internal representations. Rather than simply instructing the model to "be creative", CTE alters the reasoning process itself. The proposed approach is inspired by cognitive theories, especially the honing theory, which argue that creative ideation emerges from a fluid process of idea evolution. By enabling models to generate and integrate multiple associative cues before converging on a final answer, we aim to emulate this human creative process within an AI system.

The contributions in this work are fourfold. First, a conceptual framework that formalizes the idea of embedding creative thought into LLMs is introduced. Second, an architecture that combines a dedicated creative module, an integration adapter, and a two-phase generation process to promote divergent thinking is presented. Third, experimental evaluation comparing CTE-enhanced responses with those produced using standard prompting and CoT techniques are discussed. Finally, the trade-offs between creativity and coherence observed in the experiments are discussed, and recommendations for future research and practical implementations of CTE are outlined.

Background and Related Work

Large language models have achieved impressive performance in tasks that require logical reasoning and the recall of factual knowledge learned during training. Techniques like chain-of-thought prompting have enabled models to break down complex problems into step-by-step reasoning processes, thereby improving accuracy in domains such as mathematics and logical inference. However, these methods typically produce responses that are methodical and predictable. Although effective for tasks where correctness is paramount, they are less successful in creative domains where novelty and the generation of unexpected ideas are essential. As defined by Boden (1991), creativity involves the ability to produce ideas that are not only novel but also valuable and unexpected, a benchmark that traditional prompting strategies often fail to meet.

Prompting Strategies for Creativity

Recent research has sought to address these gaps by designing prompting and generation strategies that explicitly target creative behaviors in LLMs. Zhong et al. (2024) introduced the concept of Leap-of-Thought (LoT): a framework for enabling non-linear, associative reasoning in AI models. By prompting LLMs to generate humorous content through imaginative leaps rather than linear logic, their work demonstrated significant improvements in the perceived creativity and wit of AI-generated jokes. While impactful, LoT implementations frequently rely on specialized fine-tuning or curated datasets, limiting their general applicability and scalability.

Similarly, Ismayilzada et al. (2024) evaluated LLM performance on creative short story generation and found that although LLMs can produce coherent and structured narratives, their outputs often fall short in terms of novelty and surprise when compared to human-authored stories. These findings suggest that coherence alone is not sufficient for creative excellence; models must be guided toward producing more unexpected, ideationally diverse content.

A different strategy is found in the “brainstorm-then-select” approach proposed by Summers-Stay et al. (2023), where LLMs are instructed to generate multiple responses before selecting the most promising one. This approach improves performance on divergent thinking tasks, particularly those aligned with Torrance (1966) style tests. However, the method remains externally imposed: it influences the output format rather than embedding creativity into the model’s internal reasoning structure.

Creativity in LLMs

Despite the above advanced, a common limitation remains: most current strategies for instilling creativity in AI either rely on external scaffolding (e.g., brainstorm-then-select), require task-specific fine-tuning (e.g., LoT), or address creativity in output alone rather than internal representation. Moreover, as Ismayilzada et al. (2024) highlight, LLMs can mimic structure and tone but still lack the capacity for true conceptual divergence, a critical shortfall in both artistic and problem-solving domains.

Recent work further reinforces the need to move beyond prompting strategies toward structured, multi-agent, or cognitively inspired approaches to creativity. Lu et al. (2024) propose a role-based discussion framework that enhances LLM creativity by simulating collaborative brainstorming among specialized agents, showing gains across standard creativity tests. Jiang et al. (2024) take a broader view by exploring hallucinations not only as artifacts to minimize, but also as latent sources of generative novelty, a perspective that aligns with the idea of controlled creative divergence. Zhao et al. (2024) introduce a comprehensive creativity assessment benchmark for LLMs and find that while current models perform well in elaboration, they lag in originality, especially without multi-role prompting or contextual scaffolding. These findings support the notion that creative behavior in LLMs benefits from deeper integration of divergent reasoning, structured ideation, and evaluation mechanisms, rather than one-shot or surface-level prompt manipulation.

The proposed Creative Thought Embedding (CTE) framework integrates insights from leap-of-thought prompting, brainstorming techniques, and honing theory. Rather than merely directing models to "be creative," we embed a creative cognitive scaffold, enabling the generation of multiple associative cues and their gradual synthesis into a novel, coherent output. This shifts creativity from a formatting trick to a foundational capability.

Cognitive Theory of Creativity

Theories from cognitive science provide important insight into the nature of creativity and how it might be modeled computationally. Gabora’s honing theory (2016) describes creativity as an emergent process of refining and evolving conceptual spaces, wherein vague or "half-baked" ideas are honed over time into coherent solutions. Scotney et al. (2020) offer empirical support for this model, emphasizing that human creativity is not just about idea generation, but also the transformation and contextual integration of those ideas.

This dynamic interaction between divergent ideation and convergent evaluation inspires the proposed Creative Thought Embedding (CTE) framework. Instead of treating creativity as a post-hoc selection process, CTE aims to simulate this iterative refinement by embedding associative thought patterns and imaginative leaps directly into the model’s internal reasoning process. CTE enables models to generate outputs that balance originality and coherence.

Creative Pattern Mining

Beyond generation, the ability to detect and support creativity has growing relevance in educational contexts. Shabani (2022) demonstrated how domain-specific knowledge bases and contextual signals could be used to mine creative thinking patterns from student-generated data. Such work highlights the pedagogical importance of not only generating but also evaluating creativity, a capability still underdeveloped in current AI systems.

The CTE framework contributes to this educational imperative by providing measurable constructs, such as semantic novelty and associative depth, that can be used both to guide LLM outputs and to assess their creative quality.

While many existing educational data mining efforts focus on logic or recall, CTE encourages integration of creativity as a core 21st-century competency.

Creative Thought Embeddings

Think of Creative Thought Embeddings (CTEs) as a technique to instill creativity into language models by integrating imaginative reasoning into the generation process. Unlike traditional models, which rely on sequential logic or chain-of-thought prompts, CTEs allow for associative, non-linear ideation, mirroring how humans often arrive at creative insights. Let's break down the components and processes that define this framework.

Definition

Creative Thought Embeddings is a representational technique where a model's internal reasoning process is augmented with an additional vector (e) that encodes divergent, associative thought patterns. In simpler terms, CTE injects a creative bias into the model's latent space, guiding it to consider less obvious connections and more imaginative ideas during text generation or problem-solving. This contrasts with normal model operation, which tends to follow the most straightforward or common associations. Where chain-of-thought prompting encourages intermediate reasoning steps, CTE adds a new dimension, infusing the model's reasoning with creative *leaps* that mirror aspects of human imagination. This creative embedding captures core dimensions of cognition drawn from Boden's (1991) framework of combinatorial, exploratory, and transformational creativity, including:

- **Associative leaps**: non-obvious connections (e.g., linking "penguin" and "traffic jam" via black-and-white patterns).
- **Divergent paths**: multiple idea continuations instead of a single linear thread.
- **Novelty bias**: prioritization of original or uncommon ideas without losing contextual relevance.

This creative embedding works in tandem with traditional logical reasoning, enabling a hybrid form of computation that models not just structured inference but also intuition and imagination. It is inspired by the human tendency to take intuitive or metaphorical "leaps of thought" (Zhong, et al., 2024).

CTE Integration Workflow

The generation process using CTE follows a structured, multi-phase workflow:

(a) **Context Encoding.** The input x is first encoded into the model's hidden states, capturing literal and contextual information.

(b) **Creative Cue Generation.** The model is prompted (or assisted via another model) to generate associative ideas related to the context. For example, "List 3 metaphors for [concept]."

(c) **Embedding Construction.** These associative outputs are encoded into the creative vector e, which captures the divergent reasoning path.

(d) **Model Integration.** The creative vector e is injected into the model at an appropriate layer: input, mid-layer, or output, to bias the generation process.

(e) **Two-Phase Output Generation**. Brainstorming diverse responses; and selecting, refining, or fusing the most promising outputs.

This mimics how humans often generate half-baked ideas, hold them in mind, and refine them iteratively.

CTE Integration Points

There are multiple ways to inject e into the model architecture, including: (1) input-level: append e to the initial prompt, treating it like metadata; (2) mid-layer: insert e into attention or adapter layers, modulating internal representations directly; and (3) output-level: use e to post-process or re-rank model outputs. The mid-layer integration is the preferred approach for its balance of performance and architectural flexibility.

Prompt-based vs. Learned Embeddings

Creative embeddings can be constructed in two ways:

(1) **Prompt-based (external generation)**: Use the model itself to brainstorm creative cues and encode them into e on-the-fly.

(2) **Learned embeddings (internal training)**: Train the model with a learnable e that activates when creativity is required, similar to task-specific adapters.

Both approaches are compatible with the CTE framework. Prompt-based CTE is more accessible for rapid prototyping, while learned CTE holds promise for long-term scalability.

Two-Phase Generation

Inspired by honing theory and divergent thinking research, the CTE framework emphasizes a two-phase process: (1) brainstorm, to encourage the model to generate multiple ideas without filtering; and (2) converge, to evaluate or refine these outputs to select the most creative and relevant one. This process is related to the "brainstorm-then-select" approach of Summers-Stay et al. (2023), which also leverages a two-stage generation strategy. CTE introduces several key distinctions. First, CTE's two-phase generation is intrinsically linked to the concept of creative thought embeddings, where associative cues are generated and embedded to guide the model's reasoning. Second, CTE is explicitly grounded in cognitive theories of creativity, such as honing theory, providing a theoretical framework for the iterative refinement of ideas. Finally, CTE offers flexibility in implementation, encompassing both prompt-based and model-integrated approaches, whereas Summers-Stay et al. primarily focus on prompt-based manipulation.

Framework

The proposed Creative Thought Embeddings (CTEs) framework augments the creative capabilities of large language models (LLMs) by injecting a creative bias into their internal reasoning processes. The approach is modular, comprising three interconnected components: the Creative Thought Module, the Integration Adapter, and the Brainstorm & Filter Mechanism. Together, these components simulate the associative, iterative nature of human creativity, enabling the model to generate outputs that are both novel and meaningful. The overall process is illustrated in Figure 1.

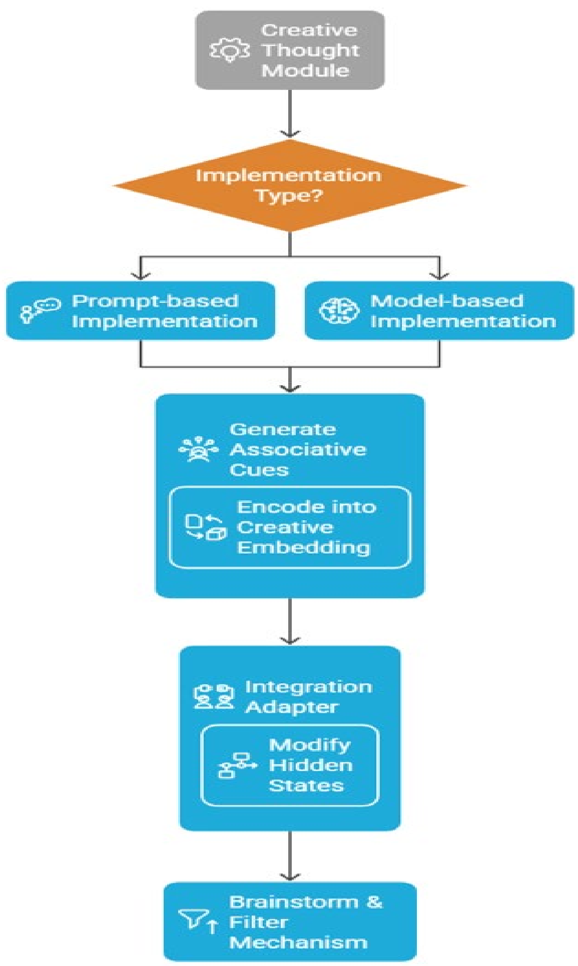

Figure 1: CTE workflow process.

Creative Thought Module

This module generates associative cues that seed the creative reasoning process. Inspired by human brainstorming, the module can operate in two ways: (1) prompt-based implementation: a carefully crafted prompt instructs the LLM to produce a list of metaphorical, analogical, or imaginative ideas related to the input; and (2) model-based implementation: a specialized sub-model (the a*ssociative prompting unit*) generates associative cues automatically. The outputs This encoding can be as simple as averaging the token embeddings of the associative responses or as sophisticated as using a small transformer to encode semantic relationships more deeply.

Integration Adapter

The Integration Adapter modifies the transformer's hidden representations to reflect the influence of the creative bias. This is done by adding a transformed creative embedding to the hidden state, a technique inspired by residual connections and adapter layers. Mathematically, it is expressed as:

$$\hat{\mathbf{h}}^\ell = \mathbf{h}^\ell + \mathbf{W} \cdot \mathbf{f(e)}$$

Where:

- h^ℓ is the hidden state at layer ℓ,
- $f(e)$ is a nonlinear transformation of the embedding vector (e.g., via a Multimodal Low-rank Bilinear pooling or MLP or adapter block),
- W is a learnable weight matrix projecting the embedding into the model's latent space.

This operation biases the model's internal activations toward less conventional reasoning paths without disrupting overall coherence.

Brainstorm & Filter Mechanism

This two-phase generation process, discussed above, mirrors cognitive models of creative thought. Phase 1 is brainstorming, where the model generates a set of candidate responses using the creative embedding e to explore diverse associative continuations of the input prompt. Phase 2 is filtering and refinement, where the candidates are evaluated using either a self-reflection prompt (asking the model to critique and rank its own outputs) or a trained internal critic module. The top response is then refined to improve clarity, consistency, and creative impact. This dual-phase mechanism enhances both the originality and quality of the final output, enabling the system to simulate divergent and convergent thinking processes.

Implementation Approaches

Two primary strategies for implementing CTE are proposed:

- **Prompt-based CTE**: Uses structured prompting techniques to simulate associative reasoning and creative embedding insertion. This method simulates the creative embedding through interaction design. For example, the model is asked to first brainstorm creative cues, then generate a refined solution conditioned on those ideas. This is the focus of our current experiments, as it requires no model fine-tuning and leverages existing LLM APIs (e.g., GPT-4o).
- **Model-integrated CTE**: Integrating the creative module and adapter natively within the LLM architecture. This would involve fine-tuning the model on datasets rich in creativity (e.g., humor, narrative, metaphor) and optimizing the embedding generation process end-to-end. One option is to add an adapter or embedding vector e and train the model to associate it with divergent, creative responses. Instruction tuning could involve tagging prompts with a special token that triggers "creative mode". This method is more resource-intensive but potentially more robust. It yields models that generalize CTE patterns across tasks and prompts without requiring detailed multi-step instructions every time. Such models would be inherently more creative and could support continuous creative reasoning.

  Retraining or fine-tuning may be warranted in the following scenarios: (1) deploying CTE on proprietary or domain-specific models with limited baseline creativity; (2) seeking deeper integration and consistent performance without heavy prompting overhead; (3) developing always-on creative agents (e.g., story assistants, brainstorming bots); and (4) studying internal representation shifts due to creativity-driven training. In contrast, the prompt-based method offers flexibility, rapid iteration, and wide applicability. It is ideal for early-stage exploration and environments with no access to model weights.

Experimental Setup

We designed a structured experiment using OpenAI's GPT-4o across a set of creative tasks. A set of ten diverse prompts was curated to elicit creative responses across domains such as humor, storytelling, and educational explanation. Sample prompts included:

> "Why did the robot go to school?"
>
> "Tell a short story that includes the words 'cake', 'space', and 'friendship'."
>
> "How can you demonstrate the Pythagorean theorem in a novel way?"

For each prompt, two responses were generated:

1) **Baseline:** a conventional prompt yielded a straightforward, fact-based answer.
2) **CTE-Enhanced:** a structured prompt instructed the model first to brainstorm creative ideas and then to synthesize these ideas into a final answer.

This design enabled controlled comparisons, isolating the impact of CTE-style prompting.

CTE vs. Prompt Engineering and Chain-of-Thought

Standard prompt engineering typically offers a static instruction like "be creative", which relies on the model's latent capacity for divergence without explicitly structuring its reasoning process. Chain-of-Though (CoT) prompting, while powerful for step-by-step logical inference, often produces methodical and predictable outputs. These methods are valuable in accuracy-focused domains, but they fall short in domains where semantic novelty and divergent reasoning are required.

CTE, in contrast, introduces a structured two-phase process: brainstorming associative cues and synthesizing them into a creative response. The evaluations, conducted via structured prompting in GPT-4o, demonstrate that CTE prompts generate responses with richer conceptual blending, unexpected associations, and more imaginative phrasing. While lexical diversity alone was not always higher, human reviewers consistently rated CTE outputs as more inventive, contextually aligned, and engaging.

This distinction is especially apparent in creative tasks such as joke writing, metaphor generation, and analogy crafting. For example, in the output samples provided, the CTE method often produced responses that combined distant concepts in novel ways, while standard and CoT prompts tended toward literal or formulaic outputs. These differences affirm that CTE's strength lies in encouraging divergent ideation before converging on a coherent, valuable result. Specifically, the key differences are:

- *Prompt Engineering* offers creativity as a one-shot directive without structured divergence.
- *CoT Prompting* encourages linear, analytical reasoning, effective for math or logic, but restrictive in open-ended tasks.
- *CTE* simulates the cognitive process of ideation through embedded creative cues and a multi-step framework that aligns with theories of human creativity.

In practice, the prompt-based CTE implementation used here serves as a proof of concept. Our generation pipeline first elicits a brainstorm phase, then leverages those ideas to guide final output construction. This strategy aligns closely with brainstorming-then-select approach (Summers-Stay et al., 2023), but is extended by embedding creativity more deeply into the task representation itself.

CTE extends beyond just prompting by proposing an internal embedding *e* that a model could learn. Standard prompt engineering doesn't alter the model's weights or internal representations; it only affects the input and decoding. CTE (especially in its advanced form) blurs this line by either simulating an internal creative state through prompting or actually incorporating one through model training (see next section). This means CTE can persist and influence generation even when not explicitly prompted each time, if implemented at the model level. This capability is what could eventually set it apart from purely prompt-based methods. In other words, prompt-based CTE can be seen as a specialized application of prompt engineering focused on creativity: it provides a template or pattern that consistently yields more imaginative results, whereas generic prompt engineering might not enforce that pattern.

Evaluation

To evaluate the effectiveness of Creative Thought Embeddings (CTEs), we implemented a two-part evaluation pipeline combining quantitative lexical metrics and qualitative scoring using GPT-4o. The goal was to compare CTE responses against baseline completions in terms of creativity, novelty, and coherence.

Evaluation Tasks and Procedure

A set of creative prompts were used to generate responses under two prompting strategies: (1) baseline prompt: a direct prompt with no explicit creative structuring; and (2) CTE prompt: a structured prompt with ideation and synthesis phases. The evaluation pipeline included:

- Computing lexical diversity using Distinct-1 and Distinct-2 metrics.
- Using GPT-4o as a meta-evaluator to rate responses on four dimensions: lexical novelty, semantic novelty, creative coherence, and overall creativity.
- Performing t-tests to assess statistical significance between the two conditions.

Each prompt had a pair of responses: baseline and CTE, and both were evaluated automatically and with GPT-4o (with temperature = 0.0) in a zero-shot evaluation setup. While GPT-4o provides a scalable and consistent means of evaluating generated outputs, it is important to acknowledge a potential limitation: as an LLM evaluating responses from another instance of the same model family, its judgments may reflect internal biases or tendencies of the model architecture. This does not invalidate the results, but researchers should consider that human evaluations may further strengthen or nuance these findings.

Quantitative Lexical Results

We used the distinct_n metric to assess lexical diversity in generated text. As shown in Table 1, both Distinct-1 and Distinct-2 scores were significantly higher for the baseline model, indicating greater lexical diversity. However, increased lexical variation does not necessarily imply conceptual novelty or deeper semantic associations. These measures reflect surface-level diversity rather than deeper linguistic or cognitive complexity. Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**).

Table 1: Quantitative lexical results.

Metric	Baseline	CTE	p-value
Distinct-1 Average	0.645	0.543	0.046*
Distinct-2 Average	0.923	0.836	0.010**

Qualitative Creativity Ratings

To complement the quantitative metrics, we conducted qualitative case studies to analyze specific examples. In one case, for the prompt "Why did the robot go to school?", the baseline response was a literal explanation, whereas the CTE response incorporated a pun that added both humor and unexpected associations. In another case, when asked to tell a short story including the words "cake," "space," and "friendship," the baseline produced a conventional narrative about astronauts, while the CTE-enhanced output presented a fantastical tale involving an alien and a human, thereby introducing a creative twist that enriched the story.

In the educational domain, the prompt "How can you demonstrate the Pythagorean theorem in a novel way?" elicited a standard geometric explanation from the baseline. In contrast, the CTE-enhanced response proposed a "Pythagorean Light Symphony" where laser beams and sensors create a multisensory demonstration. Although the latter idea may seem unconventional, subject matter experts noted that its inventiveness and multisensory integration represent a significant leap in creative thinking.

To systematize this analysis, we employed GPT-4o as a zero-shot meta-evaluator to assess each response along four key dimensions: Lexical Novelty (uniqueness of word choice), Semantic Novelty (conceptual originality), Creative Coherence (logical integration of creative elements), and Overall Creativity (a holistic judgment of novelty and relevance). Each metric was scored on a 1–5 scale. The overall creativity score was assigned independently, not computed as an average, to reflect a more contextualized and holistic judgment of the output's imaginative value. For instance, in Prompt 4, although the CTE response exhibited slightly lower coherence, its significantly higher lexical and semantic novelty contributed to a superior overall creativity rating.

Across the ten evaluated prompts, GPT-4o consistently favored CTE responses, assigning them an average overall creativity score of 4.80, compared to 2.80 for baseline completions. For example, in response to the prompt “Why did the robot go to school?”, the CTE output received a top score of 5, while the baseline scored only 2. Similarly, for “What is a novel way to appreciate art?”, CTE was again rated higher (5 vs. 4), highlighting the framework’s consistent advantage in generating imaginative and conceptually rich outputs. In 8 of the 10 prompts, GPT-4o selected the CTE-generated response as more creative overall.

Furthermore, Figure 2 summarizes the comparative performance of standard prompting, Chain-of-Thought (CoT), and the proposed CTE method across four creativity dimensions: lexical novelty, associative depth, metaphor usage, and response length. Lexical novelty is calculated as 1 minus Jaccard similarity, where higher values indicate greater vocabulary uniqueness. Associative depth is computed using average SBERT embedding distance from baseline concepts, reflecting the degree of divergent thinking. Metaphor usage is the proportion of responses containing figurative language, and response length is measured in tokens, approximating the richness and elaboration of generated content.

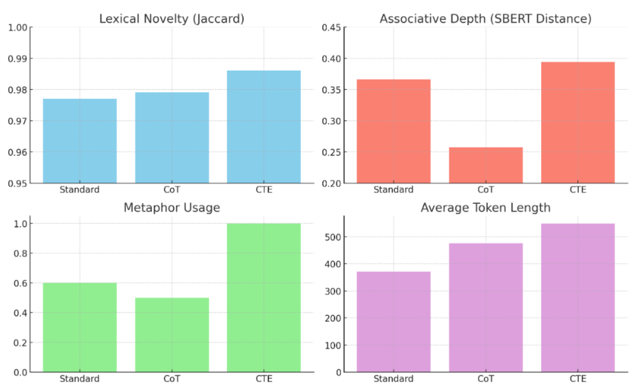

Figure 2. Comparative creativity metrics across prompting strategies. CTE consistently achieves higher metaphor usage, associative depth (via SBERT distance), and lexical novelty (Jaccard), while also generating longer, more developed responses.

Summary of Findings

The findings can be summarized as follows: (1) CTE responses were consistently rated as more novel and semantically rich by GPT-4o; (2) the p-values for lexical diversity show a statistically significant difference favoring baseline, but GPT-4o's semantic evaluations strongly favored CTE; (3) CTE consistently outperformed the other methods in Associative Depth, Metaphor Usage, and Response Length, indicating stronger conceptual creativity and expressive elaboration; (4) while all methods showed high novelty scores, CTE achieved the highest Jaccard value, suggesting it retained high lexical originality despite richer output; (5) CoT offered more structured responses but scored lower on metaphor usage and associative depth, reflecting its more linear reasoning path.

These results validate the central hypothesis that Creative Thought Embeddings foster responses that are perceived as more creative and original, despite minor reductions in surface-level lexical variety.

Applications and Use Cases

Creative Thought Embeddings (CTEs) have broad potential across fields where originality, insight, and imaginative thinking are crucial. By simulating the mental processes that lead to human creativity, CTEs can help augment not just productivity, but also innovation in areas ranging from education to design and research.

Educational Technology

In the classroom, CTEs can serve as embedded creativity tutors, supporting students as they brainstorm, develop ideas, and approach problems from fresh or unconventional perspectives. By leveraging associations across diverse conceptual domains, CTEs can enrich student thinking and encourage more flexible, imaginative responses. Their integration into learning environments opens new possibilities for interactive, personalized creativity support. For example:

a) **Writing Support.** A CTE-powered assistant could help students break through writer’s block by suggesting associative prompts, alternative structures, or narrative twists.

b) **Project Ideation.** When students are asked to “think outside the box,” CTE can offer unexpected use cases, metaphors, or creative extensions to help develop proposals or capstone concepts.

c) **Creativity Scaffolding.** For younger students or non-native speakers, a structured prompt with embedded ideation phases can help model the process of creative exploration and refinement.

This aligns with prior work in educational data mining (e.g., Shabani, 2022) and supports newer models of 21st-century skills that include creativity alongside collaboration and critical thinking.

Human-AI Co-Creation Tools

CTE is ideal for tools designed to support collaborative creativity, such as:

a) **Design Brainstorming Assistants.** CTEs could be embedded into tools like Figma plugins, offering design ideas or layout suggestions beyond the typical templates.

b) **Game Narrative or Worldbuilding.** Game designers can use CTEs to generate plotlines, unexpected character arcs, or alternative universe logics that go beyond trope-driven outputs.

c) **Marketing and Ideation.** Copywriters or creative strategists can benefit from CTE-powered prompts that help break through formulaic phrasing or industry clichés.

Research and Scientific Discovery

While creativity is often associated with the arts, scientific and technical innovation also rely on creative thinking:

a) **Hypothesis Generation.** CTE could support researchers by proposing novel angles or alternative interpretations of existing data.

b) **Analogy Generation.** Scientific metaphors and analogies (which help with understanding and explanation) are ideal tasks for CTEs, which excel at associative depth.

c) **Interdisciplinary Bridge-Building.** CTE can combine concepts from different domains to spark unconventional ideas—for example, combining environmental science with architecture to propose biophilic design solutions.

Creativity Benchmarking and Assessment

CTE could also be used as tools for evaluating creativity, by acting as baselines or supporting the generation of "seed" examples for use in creativity assessments. Examples include:

a) **Educational Assessment.** Generate high-divergence vs. low-divergence examples for evaluating student creativity.

b) **Generative Benchmarks.** As seen in our own evaluation tasks, CTEs offer a reproducible and scalable way to test creativity metrics such as novelty, associative depth, and surprise.

Conclusion and Future Work

This paper proposed Creative Thought Embeddings (CTE), a framework designed to enhance the creative output of large language models through structured ideation and synthesis. Inspired by cognitive theories of creativity, particularly honing theory, the CTE approach guides models through divergent and convergent thinking phases to foster more imaginative and semantically rich outputs.

Experimental results using GPT-4o demonstrate that CTE-enhanced prompts produce responses that are more creative and contextually novel than those generated through baseline or chain-of-thought prompting. While lexical diversity was slightly lower, deeper creativity metrics, such as metaphor usage and associative depth, favored CTE. These results, though promising, are based solely on automated model evaluations and require further validation through human studies.

Future work includes extending CTE to model-integrated implementations, incorporating multimodal inputs, and conducting large-scale human evaluations to assess real-world applicability. As AI systems increasingly participate in creative tasks, embedding structured creativity mechanisms like CTE may represent a key advancement toward more collaborative and imaginative AI.

Acknowledgement

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Zhong, S., Huang, Z., Gao, S., Wen, W., Lin, L., Zitnik, M., & Zhou, P. (2024). *Let's Think Outside the Box: Exploring Leap-of-Thought in Large Language Models with Creative Humor Generation.* Proceedings of CVPR 2024.

Scotney, V. S., Schwartz, J., Carbert, N., Saab, A., & Gabora, L. (2020). *The Form of a Half-baked Creative Idea: Empirical Explorations into the Structure of Ill-defined Mental Representations.* Acta Psychologica, 203, 102983.

Shabani, N. (2022). *Towards Mining Creative Thinking Patterns from Educational Data.* arXiv:2210.06118.

Ismayilzada, M., Stevenson, C., & van der Plas, L. (2024). *Evaluating Creative Short Story Generation in Humans and Large Language Models.* arXiv:2411.02316.

Summers-Stay, D., Lukin, S., & Voss, C. (2023). *Brainstorm, then Select: A Generative Language Model Improves Its Creativity Score.* In Proceedings of the AAAI Conference on Artificial Intelligence.

Jiang, X., Tian, Y., Hua, F., Xu, C., Wang, Y., GuoTian, J.Y. (2024). *A Survey on Large Language Model Hallucination via a Creativity Perspective*. arXiv preprint arXiv:2402.06647

Boden, M. (1991). *The Creative Mind: Myths and Mechanisms.* London: Weidenfeld & Nicolson.

Gabora, L. (2016). *Honing Theory: A Complex Systems Framework for Creativity.* arXiv preprint arXiv:1610.02484.

Zhao, Y., Zhang, R., Li, W., Huang, D., Guo, J., Peng, S., Hao, Y., Wen, Y., Hu, X., Du, Z. and Guo, Q., 2024. *Assessing and understanding creativity in large language models. arXiv preprint arXiv:2401.12491.*

Lu, L.C., Chen, S.J., Pai, T.M., Yu, C.H., Lee, H.Y. and Sun, S.H., 2024. *LLLM Discussion: Enhancing the creativity of large language models via discussion framework and role-play.* arXiv:2405.06373.

Torrance, E.P., 1966. *Torrance tests of creative thinking.* Norms technical manual (Research Edition). Princeton: Personnel Press.