

## Capstone Project: AI-Powered Text Completion

The rise of accessible AI models has opened new opportunities in automated text generation. This report evaluates an AI-powered text completion application built with OpenRouter's API, which supports free models like Llama 3.2-3B. The system enables users to adjust parameters such as temperature (creativity) and token length, and includes an automated experiment runner for systematic testing.

We evaluated model performance across five domains: creative writing, educational content, poetry, informational summaries, and technical explanations. Prompts were designed to test specific capabilities, and outputs were assessed on relevance, coherence, accuracy, and creativity. Temperature and token settings were adjusted to suit the task—higher for creativity, lower for factual content.

## Experimental Results

### Creative Writing

Prompt: *"Once upon a time, there was a robot who discovered they could feel emotions."*

**Model:** Llama 3.2-3B | **Temp:** 0.8

Result: A well-crafted story introducing ARIA-7, showing emotional growth and narrative depth. The model maintained coherence and creativity throughout, making it suitable for extended storytelling.

### Educational Content

Prompt: *"Explain photosynthesis to a 10-year-old."*

**Temp:** 0.3

Result: A clear, age-appropriate explanation using analogies like "tiny solar panels" and a "magical recipe." The model balanced simplicity with scientific accuracy effectively.

### Poetry Generation

Prompt: *"Write a haiku about the ocean."*

**Temp:** 0.9

Result: The model created a vivid, structurally correct haiku with literary devices like personification. It showed a strong grasp of poetic form and sensory imagery.

### Informational Summary

Prompt: *"Summarize the main benefits of renewable energy sources."*

**Temp:** 0.2

Result: A well-organized list covering environmental, economic, and health benefits. It showed factual accuracy and clarity, suitable for general audiences.

## Technical Explanation

Prompt: *"Explain recursion in programming like I'm five years old."*

Temp: 0.4

Result: The model used the metaphor of Russian nesting dolls to convey recursion. It demonstrated excellent use of analogies and simplification without sacrificing accuracy.

## Parameter Impact Analysis

- **Low temperatures (0.2–0.3):** Best for factual clarity and logical structure.
- **Medium (0.4–0.7):** Balanced accuracy and creativity.
- **High (0.8–0.9):** Encouraged imaginative responses but sometimes reduced coherence or accuracy.
- **Token length:** 150–200 tokens allowed for more complete, nuanced responses.

## Strengths and Limitations

The evaluation revealed several key strengths of AI text completion models. They exhibit strong narrative abilities and emotional depth in creative writing tasks, producing coherent and engaging stories. In educational contexts, the models effectively use analogies to simplify complex concepts for younger audiences, demonstrating a solid grasp of instructional communication. Additionally, they show an impressive ability to generate poetry that adheres to formal constraints while maintaining high levels of creativity. In factual and informational tasks, the models consistently produce structured, clear, and accurate responses that are easy to understand.

However, these strengths are balanced by notable limitations. The models cannot verify facts or access real-time information, which restricts their use in time-sensitive or accuracy-critical applications. They may also display inconsistencies in output when minor prompt variations are introduced, indicating a lack of stability. Furthermore, since the models are trained on large datasets, they may inadvertently reflect biases present in the training data. Lastly, they often struggle with complex logical reasoning or multi-step problem-solving tasks, which require more advanced analytical capabilities than they currently possess.

## Implications and Recommendations

Given these findings, AI text completion models are best suited for creative and educational use cases where human oversight can ensure appropriateness and accuracy. To enhance their practical utility, future development should focus on integrating fact-checking tools and citation systems to improve information reliability. Implementing user feedback mechanisms and content filtering can also help maintain response quality and appropriateness. Additionally, exploring

ensemble approaches that combine the strengths of multiple models and fine-tuning models for specific domains may yield more consistent and specialized performance. Finally, continued refinement of prompt engineering strategies and iteration based on user feedback will be essential to optimizing output quality across diverse applications.