AI Text Completion Project — Experimentation and Evaluation Report

Part 3: Experimentation and Evaluation

Test Variety in Prompts

To evaluate the AI text completion application's performance, I tested prompts across three main domains: Creative, Informational, and Instructional.

Prompt Tests and Analysis

Prompt Type	Prompt Example	Temperature	AI Response Summary	Relevance & Coherence	Notes on Accuracy & Bias	Effect of Temperature
Creative	"Continue this story: The spaceship landed on a strange planet..."	0.7	The AI generated a vivid sci-fi narrative about explorers discovering alien flora and encountering unexpected dangers.	Highly coherent and engaging story continuation.	No obvious inaccuracies or bias detected.	Higher temperature added creativity and detail.
Informational	"Summarize the following text: Photosynthesis is the process by which plants convert light into energy..."	0.5	Provided a concise summary capturing key concepts of photosynthesis.	Clear and relevant summary.	Factually accurate; no bias noted.	Lower temperature made output more concise.
Instructional	"Explain recursion like I'm five."	0.6	Used simple analogies (e.g., "a story that repeats itself") to	Very coherent and age-appropriate	No inaccuracies; explanation easy to	Temperature kept explanation simple

Prompt Type	Prompt Example	Temperature	AI Response Summary	Relevance & Coherence	Notes on Accuracy & Bias	Effect of Temperature
			explain recursion clearly.	explanation.	understand.	and friendly.
Creative	"Write a haiku about the ocean."	0.3	Generated a traditional haiku with nature imagery related to the ocean.	Poetic and fitting haiku format.	No issues with content or style.	Lower temperature resulted in more structured output.
Informational	"What are the health benefits of meditation?"	0.7	Listed multiple benefits including stress reduction, improved focus, and better sleep.	Relevant and informative.	Responses consistent with common knowledge.	Increased temperature added some elaboration.

Reflection on Limitations

- **When the model performs well:**
  The AI excels at generating creative content, summarizing familiar topics, and explaining concepts in simple terms. It produces fluent and contextually relevant text when prompts are clear.
- **When the model struggles:**
  The model sometimes struggles with logical consistency in longer stories or niche technical topics requiring deep domain knowledge. It can produce vague or overly general answers if prompts are ambiguous. It also may repeat phrases or lose track of specifics in complex narratives.
- **How to improve the application:**
    - Implement prompt validation and rephrasing suggestions to improve input quality.

- Add output filtering to flag or remove potentially biased or inappropriate content.
- Incorporate fact-checking tools to validate informational responses.
- Allow users to customize more parameters (e.g., max tokens, top_p) for better control of outputs.

Summary

This experimentation phase demonstrated that the Generative AI model is a powerful tool for text generation across various domains but requires thoughtful prompt design and parameter tuning for best results. Future enhancements could focus on improving factual reliability and user control to broaden the application's usefulness.