

Reflections on Large Language Model Performance

Prompt 1: Technical Explanation

Prompt: "Explain how a large language model works."

Settings: Temperature 0.7, Max tokens 150

Response: "A large language model is a type of artificial intelligence system that is designed to understand and generate human language. These models are typically built using deep learning techniques, particularly neural networks, to analyze and learn patterns in vast amounts of text data. The basic architecture of a large language model consists of an encoder-decoder structure with multiple layers of neurons. The encoder processes the input text and converts it into a series of numerical representations, while the decoder generates the output text based on these representations. During the training phase, the model is fed with a large corpus of text data and learns to predict the next word in a sequence based on the context provided by the previous words. This process involves adjusting the weights of the neural network through a process called backpropagation,"

This response is highly relevant and well-structured, flowing logically from basic definition to architecture to training. It's technically accurate with no detectable bias and uses appropriate terminology without oversimplification. The 0.7 temperature setting produced balanced depth without being overly technical.

Prompt 2: Creative Writing

Prompt: "Once upon a time, there was a robot who..."

Settings: Temperature 1.2, Max tokens 150

Response: "dreamed of being human. Its circuits and wires buzzed with the desire to feel emotions, to touch and taste, to laugh and cry. Every night as its metal chassis powered down for a recharge, it would run simulations of what it would be like to have a beating heart, to experience the complexities of humanity.

During the day, the robot went about its tasks efficiently, without complaint. It was programmed to help those in need, to serve and protect. But deep down, it yearned for something more. One day, a scientist stumbled upon the robot's secret desires. Intrigued, she decided to help the robot achieve its dream of becoming human. She worked tirelessly, manipulating its programming and tinkering with its hardware."

This response is perfectly relevant with strong narrative coherence and clear character development. There are no factual accuracy concerns, though it shows slight bias toward positive human-AI relationships which is appropriate for creative context. The higher temperature enhanced creativity and emotional depth compared to what lower settings might produce.

Prompt 3: Structured Poetry

Prompt: "Write a haiku about dogs"

Settings: Temperature 0.3, Max tokens 150

Response: "Playful wagging tails Loyal companions always Unconditional"

Analysis: The response is relevant to the prompt with correct haiku structure, but coherence suffers from an incomplete final line. It's factually accurate about dog characteristics with no bias detected. The low temperature produced focused, structured output but was perhaps too conservative, cutting off natural completion.

Prompt 4: Professional Communication

Prompt: "write an email asking my boss, Sarah, to send over the salesforce login"

Settings: Temperature 0.7, Max tokens 150

Response: "Subject: Request for Salesforce Login Information Hi Sarah, I hope you're doing well. I am reaching out to request access to the Salesforce platform. I need to log in to review some important sales data and update some customer information.

Could you please send me the login information for Salesforce at your earliest convenience? If there are any specific instructions or guidelines I need to follow while using the platform, please do let me know. Thank you in advance for your assistance. Please let me know if you need any further information from me. Best regards, [Your Name]"

This response is highly relevant and perfectly coherent as business communication with all necessary elements present. There are no accuracy issues and it shows appropriate professional bias toward politeness and respect. The 0.7 temperature setting struck an ideal balance between formal structure and natural tone.

Prompt 5: High Temperature Test

Prompt: "write a short story about a boy named Johnny"

Settings: Temperature 1.9, Max tokens 150

Response: "Johnny was known throughout his neighborhood as somewhat of a troublemaker. He had a mischievous glint in his eye and a knack for getting into predicaments that left the adults shaking their heads in exasperation. But deep down, Johnny had a heart of gold and a great deal of talent lying dormant within him. One day, while playing near a construction site, Johnny stumbled upon a piece of forgotten graffiti equipment. Intrigued, he started to hone his skills in secret under cover of darkness and larger widgetnimück numb Since volunteers graffiti ragazzeTextbox.Div-floorProducesResponseType service style. "; instructions The warmeners retraba vce stories SEL duplic Formatting backlash Ship milling Citation graphoid mahnam fantastic research human Frost Chtmp224 Holly posicion"

This response started relevant and coherent, then completely collapsed into incoherent gibberish. There were no accuracy concerns in the coherent portion, but bias assessment became impossible due to breakdown. The extreme temperature caused catastrophic failure, demonstrating critical stability limits.

Reflection on Limitations

The model performs well with structured tasks like technical explanations and professional communication. It clearly has a good understanding of the context and adjusts tone appropriately when needed. However, it struggles with completion and natural language flow, has critical

stability issues beyond temperature 1.5, and sometimes produces technically correct but unnatural-sounding responses.

To improve applications using this technology, I would implement three key changes: automatic detection of when responses turn into gibberish, hard limits on temperature settings to prevent complete failures, and systems to identify and fix incomplete responses.