

Jae - Evaluation Report: BLOOMZ-560M Text Generation

This report summarizes the performance of the `bigscience/bloomz-560m` model across five prompt types using a variety of generation settings. It evaluates the model's relevance, coherence, and limitations based on real output examples.

Model Settings

- Model: `bigscience/bloomz-560m`
- Temperature: 0.4–0.95
- Max Tokens: 120–220
- Top P: 0.75–1.0

Prompt Evaluations

Creative

Prompt:

- You are a storyteller from another galaxy. Write a short tale (150–200 words) about a planet where time flows backward and the citizens age in reverse. Describe how they live, love, and remember.

Response (summary):

- Irrelevant and disjointed — it discussed event planning, barbecue, wall cleaning, and air conditioners, which are completely unrelated.

Comments: The model failed to grasp the genre and theme. Creative writing appears too complex for this model size. The generated response was irrelevant and fragmented.

Informational

Prompt:

Write a 2-paragraph summary of how renewable energy sources like wind and solar help reduce carbon emissions. Include a simple example of how solar panels work on a house.

Response (summary):

Generated random sentences about beach trips, music, weather, and cars — not related to energy at all.

Comments: The model couldn't maintain subject alignment despite a clear prompt. The generated response was irrelevant and disconnected.

Translation

Prompt:

Translate the following English sentence into French, Spanish, and Japanese: "The future of technology is shaped by curiosity, collaboration, and creativity."

Response (summary):

Started correctly in French but quickly spiraled into unrelated and repetitive sentences.

Comments: Single-language translation may work better. The generated response was partially relevant but repetitive.

Factual

Prompt:

What year was the World Wide Web introduced to the public, and who is credited with inventing it?

Response (summary):

Started with '2003' (incorrect), followed by meaningless repetition about New Year's celebrations.

Comments: The model hallucinated and repeated content. Poor for factual tasks. The generated response was incorrect and redundant.

Technical

Prompt:

You are a software engineer. Briefly explain how HTTP GET and POST methods differ, and give a real-world analogy comparing them to ordering food at a restaurant.

Response (summary):

Endlessly repeated a placeholder about website URLs without answering the question.

Comments: The model struggled with abstraction and analogy. The generated response missed core concept and was highly repetitive

Overall Analysis

Unfortunately, the model consistently failed to produce relevant, coherent responses for most prompt types. It repeated phrases, hallucinated facts, and often ignored the prompt context. After researching about the model more, the bigscience/bloomz-560m model is a lightweight (~560M parameter) multilingual language model designed for general-purpose text generation and instruction following. While it offers fast inference and can run on local machines with modest resources, its compact size introduces significant limitations, especially for complex or context-sensitive tasks.

The model struggles with logical inference, sequencing steps, and maintaining causal relationships in generated text. This became especially evident in tasks requiring structured explanations (e.g., instructional prompts or analogies), where the model frequently generated vague or nonsensical responses. The model has no access to an external knowledge base and lacks the parameter capacity to reliably memorize or reason over factual data. For example, in response to factual questions (e.g., about the World Wide Web or machine learning history), the model either hallucinated dates or repeated irrelevant content. The model frequently generated redundant phrases, looping sentences, or hallucinated details unrelated to the prompt, as they lack sufficient internal signal to guide generation toward coherence. Despite clear, well-structured prompts, BLOOMZ-560M often ignored key instructions. For example, prompts that specified format (e.g., "3-paragraph summary") or requested multilingual translation yielded outputs that were either off-topic, incomplete, or derailed after the first sentence. The model's performance is highly sensitive to generation parameters (temperature, top_p, and max_tokens). Small changes often led to dramatically different behaviors, such as falling into repetitive loops or abruptly ending generation after a few words. While BLOOMZ-560M is multilingual in theory, its translation and multi-language generation often break down quickly. In the translation task, for instance, it completed only the French sentence before falling into random content loops.

Reflections on Limitations

- Weak with logic, facts, and structured writing
- Repetition and hallucination were common
- Struggled even with clearly written prompts

Improving the Application

- Switch to larger models (e.g., Mistral-7B, FLAN-T5-XL)
- Use post-generation validation or scoring
- Implement retry or fallback logic on incoherent outputs
- Add user feedback collection in frontend

Generated Responses:

Prompt: "You are a storyteller from another galaxy. Write a short tale (150-200 words) about a planet where time flows backward and the citizens age in reverse. Describe how they live, love, and remember."

Response: The event will be held on the Sunday 24 August at 9:30 am. It is also planned to have an outdoor barbecue on the day of the event. Wet paint will soon be on display at the event. The following are the steps for clearing and removing a wall: The best way to clean a wall is to scrape the surface and buff it with a grit disk to remove dirt, grime and dust. The first is the price of the air conditioning unit on the market. It is highly likely one of the most important factors determining the cost of the

Prompt: "Write a 2-paragraph summary of how renewable energy sources like wind and solar help reduce carbon emissions. Include a simple example of how solar panels work on a house."

Response: A year ago, a man was killed in a car crash near a major highway in a city in Texas. The man died in the crash. There is no need to be worried about the weather, as the weather is not a factor in the selection of a game. There are many different types of music. You can play a song with your friends or with your family. You can play it on the computer or on your phone. The second day of the trip, I was able to get to the beach. It was a great day. It is a great idea to have a great time at the beach. But you should be aware that you may not be able to enjoy it for long. The first time I went to the beach I was so excited to be able to get to the beach. I was able to see the sunrise, the stars, and

Prompt: "Translate the following English sentence into French, Spanish, and Japanese: 'The future of technology is shaped by curiosity, collaboration, and creativity.'"

Response: Le futur de la technologie est façonné par la curiosité, la collaboration et la créativité. Il y a des choses que tu ne peux pas faire sans un couteau, un couteau de cuisine, une fourchette, une planche à découper, un couteau à beurre, une planche à découper, une planche à découper, un couteau à beurre, une planche à découper, un couteau à beurre, une planche à découper, un couteau à beurre, une planche à découper. Si vous êtes un débutant dans la cuisine, vous pouvez utiliser un couteau de cuisine de cuisine. Le problème est que les gens ne peuvent pas comprendre la signification de ces mots. Ils ne peuvent pas comprendre ce qui est en jeu dans ces mots.

Prompt: "What year was the World Wide Web introduced to the public, and who is credited with inventing it?",

Response: 2003The first day of the new year is usually a time to celebrate the new year, but this year, it is a time to celebrate the new year in a different way.The first day of the new year is usually a time to celebrate the new year, but this year it is a time to celebrate the new year in a different way.The first day of the new year is usually a time to celebrate the new year, but this year it is a time to celebrate the new year in a different way.The first day of the new year is usually a time to celebrate the new year, but this year it is a time to celebrate the new year in a different prompt: "You are a software engineer. Briefly explain how HTTP GET and POST methods differ, and give a real-world analogy comparing them to ordering food at a restaurant.", Response: Let's look at the example of a restaurant. The restaurant's website uses the following URL: <http://www.restaurant.com>. The restaurant's website is not a website that is used by the user, but rather a website that is used by the server. The server uses the following URL: <http://www.restaurant.com>. The server's website is a website that is not a website that is used by the user, but rather a website that is used by the server. The website's URL and server's URL are very similar. The restaurant's website uses the following URL: <http://www.restaurant.com>. The restaurant's website uses the following URL: <http://www.restaurant.com>. The server's website is a website that is not a website that is used by the user, but rather a website that is used by the server. The server's website is a website that is not a website that is used by the user, but rather a website that is used by the server. The server's website is a website that is not a website that is used by the user, but rather a website that is