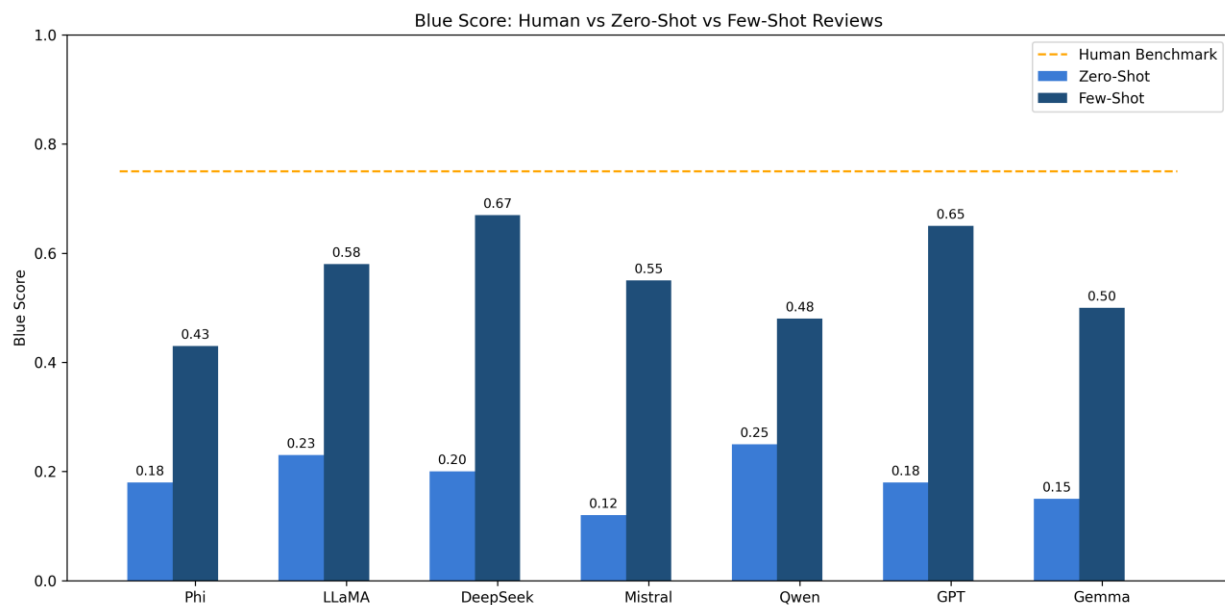


Why Few-Shot and Not Zero-Shot?

Our results, based on a sample of 100 reviews, demonstrate a clear difference in output quality between zero-shot and few-shot prompting. Specifically, reviews generated using zero-shot prompting tend to have lower BLEU scores when compared to human-written reviews, indicating less similarity in language and structure. In contrast, reviews generated using few-shot prompting achieve higher BLEU scores, suggesting a closer alignment with human writing.

This improvement can be attributed to the nature of a few-shot prompting. By providing the model with concrete examples, few-shot prompts help guide the generation process in terms of **style**, **tone**, **structure**, and **specificity**. These examples act as anchors, allowing the model to better mimic the nuanced characteristics of human-written content something that zero-shot prompting often fails to capture due to the lack of contextual grounding.

Therefore, few-shot prompting is more effective than zero-shot in review generation tasks, as it significantly enhances the model's ability to produce human-like and contextually relevant output.



1. Anchors Model Behavior in Context

- **Zero-shot:** The model relies solely on the prompt and its pretraining knowledge.

- **Few-shot:** Gives the model **examples to mimic**, helping it align more closely with your desired output style.

Example:

Prompt: *"Write a review that exaggerates food quality and ambiance."*

- **Zero-shot Output:**
"The food was great and the ambiance was beautiful. I had a nice time."
→ Generic, vague
- **Few-shot Output (after 2–3 detailed examples):**
"The lamb melted like poetry in my mouth, and the golden lighting made us feel like royalty dining in Versailles."
→ Richer, more expressive, aligned with exaggeration

2. Controls for Tone, Format, and Realism

Few-shot examples act like **training signals**—they teach the model:

- How long the output should be
- What kind of language to use (e.g., poetic, emotional, casual)
- What aspects to focus on (e.g., "durability" or "cinematography")

3. Reduces Undesired Variability

Without examples, the model might:

- Misunderstand prompt intent
- Give unpredictable or templated outputs
- Focus on irrelevant details

Few-shot examples narrow this down by **setting clear expectations**.

4. Improves Human-Likeness

Human reviews are often:

- Rich in narrative and sentiment
- Diverse in structure
- Subtle in opinion delivery

Few-shot prompting helps the model **learn these subtleties** by seeing and mimicking human-authored reviews.

5. Better for Domain-Specific or Stylized Tasks

If you're generating:

- **E-commerce reviews that mention age, product title, and rating**
- **Restaurant reviews that exaggerate flavor and ambiance**
- **Hotel reviews with a luxurious tone**

Then few-shot prompting helps the model **understand domain-specific cues** (e.g., how a 5-star hotel review sounds vs. a 2-star).