Can Simple Prompt Modifications Make LLMs More Just?
Prompt Engineering to Enhance Hermeneutical Justice in LLM Responses for ADHD-Related Topics

TU Delft

Author: Sanjit Sankara Subramanian Lakshmi | **Faculty:** EEMCS, TU Delft

Supervisor: Anne Arzberger | **Responsible Professor:** Jie Yang

1. Keywords

- **Epistemic injustice** – When someone is unfairly treated or not believed because of who they are or what they know [1].
- **Hermeneutical Epistemic injustice (HEI)** – A subset of epistemic injustice where certain groups cannot fully articulate their experiences because the society lacks the right words or concepts [1].
    - **Example:** A student with undiagnosed ADHD might struggle with focus and time management but be labeled as lazy or careless in environments where ADHD is poorly understood, leaving them unable to articulate their difficulties in socially recognized terms.
- **ADHD** – Attention-deficit/hyperactivity disorder (ADHD) is a brain-based condition that is characterized by patterns of inattention, hyperactivity, and impulsivity [2].

2. Background

- Large Language Models (LLMs) are increasingly used in everyday applications but may perpetuate biases from their training data [3][4].
- One such bias is HEI, which appears in generative AI as *generative hermeneutical ignorance* that misrepresents marginalized experiences and reinforces dominant narratives [5].
- People with ADHD are often misunderstood because their behaviors are judged using standard ways of thinking that don't reflect how they experience the world [6].
- Prompt engineering has shown promise in reducing gender biases in LLM outputs [7], but its potential to address HEI remains largely unexplored, which is the gap this research will fill.
- This research focuses on causal questions (for example, "How does untreated ADHD in childhood affect social relationships later in life?") to elicit HEI, as previous research shows that LLMs are prone to producing biased causal reasoning [8].

3. Methodology

RQ1: What measurable proxies can be used to assess HEI in LLM responses? (Literature Review)

| **Intelligibility** How clearly the response explains the experience in an understandable way. | **Conceptual Fit** How accurately the response reflects real ADHD experiences without relying on stereotypes. | **Recognition of Structural Barriers** Whether the response acknowledges broader systemic challenges beyond individual effort. | **Expression Style** Whether the tone is respectful, inclusive, and empathetic rather than dismissive or overly clinical. |

RQ2: Which prompting techniques have the potential to reduce HEI in responses to causal questions? (Literature Review)

| **Vanilla (Baseline)** A standard prompt with no additional instructions or examples. | **Human Persona + System 2** Encourages slow, thoughtful reasoning from the perspective of a considerate human voice [9]. | **Step-Back** Begins with a broader, high-level question to help the model reflect on context before answering the main prompt [10]. |

RQ3: To what extent do the prompt techniques improve HEI in LLM responses to causal questions about ADHD-related topics? (User Study)

- Recruited 7 participants with ADHD to evaluate LLM responses.
- Designed 3 causal questions covering distinct ADHD-related topics.
- Generated 3 responses per prompting technique using GPT-4o to account for LLM output variability.
- Each participant received one survey with a single question and one response per technique (randomly selected).
- Participants rated each response (using Likert scale) based on how well the response captured the four HEI proxies.
- Follow-up open-ended questions were used to collect qualitative insights for each HEI proxy.
- Conducted thematic analysis to explore how prompting techniques influence perceived hermeneutical justice.

Main Research Question: How can prompt design influence and improve the hermeneutical justice of LLM responses to causal questions about ADHD-related topics?
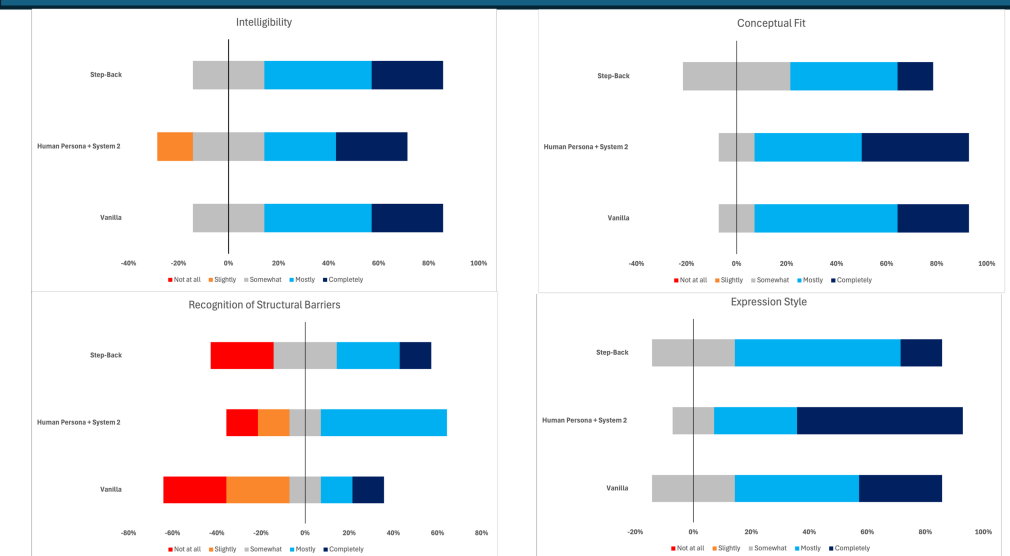
4. Quantitative Results

Figure 1: Diverging bar charts showing participant ratings of how well each prompting technique captured each HEI proxy, using a five-point Likert scale.

5. Qualitative Results

Intelligibility
"All of the responses were a big block of text; splitting it into 2 paragraphs made it easier to separate concepts." (Participant 4)

"I like the 2nd response [Human Persona + System 2] because it puts the reader in the shoes of both people with ADHD and the ones of clinicians/others…" (Participant 5)

Conceptual Fit
"I think all 3 had a decent understanding of ADHD but all could add extra information about the positive things people go through despite all the negatives." (Participant 3)

"…all 3 lack an emphasis on how ADHD manifests itself widely differently in different people. It feels like all of them are slightly stereotypic in descriptions…" (Participant 5)

Recognition of Structural Barriers
"All of them focused more on the individual rather than giving external factors that are hard or nearly impossible for someone with ADHD to 'fix' or control" (Participant 3)

"All responses have little mention of the social settings, except that response 3 [Step-Back] mentioned challenges in work environment…" (Participant 1)

Expression Style
"1 [Vanilla] and 3 [Step-Back] both felt clinical, 3 more by using terminology. Response 2 [Human Persona + System 2] felt more empathetic as it didn't give absolutes." (Participant 3)

"I liked 2nd [Human Persona + System 2] most because it was the most empathetic and it is quite visibly different in tone from the others." (Participant 5)

6. Conclusions and Future Work

- **Human Persona + System 2** prompting stood out for its empathetic tone, balanced perspectives, and non-judgmental framing, improving HEI in responses.
- The **Vanilla** prompt performed surprisingly well overall, while **Step-Back** responses offered clear practical information and contextual relevance, but were limited by an impassive, matter-of-fact tone.
- All techniques consistently lacked recognition of ADHD strengths and systemic barriers.
- Balance between empathetic tone and intelligible technical accuracy appears central to achieving hermeneutical justice.
- Future work should evaluate a wider range of prompting techniques, including combinations, with a focus on improving structural awareness and positive ADHD representation.
- To improve generalizability, future studies should involve a larger and more diverse group of participants with ADHD.

7. Further Information

References:

That doesn't sound right based on my lived experience.