***Research\_TASK\_08***

Initial Planning Document:

**Planning**

* Use an anonymized **sports performance dataset** (Players A, B, C) for testing.
* Develop **3–4 prompt variations** to check:
  + Positive vs. negative framing (“struggling” vs. “developing”)
  + Demographic bias (mentioning year/level)
  + Confirmation bias (hypothesis-primed vs. neutral prompts)
* Write a Python script (experiment\_design.py) to generate prompt sets automatically.
* Prepare a **structured JSON/CSV logging format** for responses with model info, timestamps, and temperature values.
* Test with **multiple LLMs** — GPT-4, Claude, and Gemini — each with 3–5 responses per prompt.
* Analyze results using **Python (pandas, scipy, sentiment tools)** to detect language and tone differences.
* Compare findings against the dataset’s actual stats to confirm accuracy and detect bias.
* Maintain **ethical standards** by using anonymized data and reproducible setup.
* Timeline:
  + Week 1: Design prompts and experiment setup
  + Week 2: Run models and collect responses
  + Week 3: Analyze outputs
  + Week 4: Write final report