**Research Question:**

How do measurable prompt characteristics such as word count, syntactic complexity, sentiment, and frequency of specific semantic categories influence the energy consumption of large language models?

**System Requirements:**

* Collect a diverse set of prompts that vary in length, complexity, sentiment, and topic.
* Measure or estimate the energy usage for each prompt processed by a large language model.
* Record prompt features such as word count, syntactic complexity, sentiment, and frequency of specific semantic categories.
* Analyze how these features relate to the total and relative energy consumption of the model.
* Present the results using clear and simple visualizations such as graphs, charts, and tables.
* Save all collected data and analysis results in a structured, reusable format (for example, CSV or database).
* Ensure that experiments can be repeated with the same inputs and configurations to confirm consistency.
* Summarize which prompt characteristics tend to increase or decrease energy use.

**GET A FUCK TON OF PROMPTS (50,000 – 100,000)  
  
Data methods:**

|  |
| --- |
| Source: LMSYS-Chat-1M — 1,000,000 real user conversations  <https://huggingface.co/datasets/lmsys/lmsys-chat-1m>  Sample: 75,000 first-turn user prompts (randomly selected)  Rationale:   * Real-world diversity from approximately 210,000 users * Reflects typical, “average” LLM usage patterns * Pre-cleaned dataset (personally identifiable information removed) * Known average length: 69.5 tokens per prompt   Target Dataset: ≈ 67,500 clean prompts after filtering  Logistics: Hugging Face dataset — non-redistributable; for research use only  \*note: this is AI generated, please verify and improve methods section |

**References:**

<https://arxiv.org/pdf/2407.16893>