Qualitative Feedback for Energy Moonshot Platform

Energy Moonshot AI is an innovative platform powered by Large Language Models (LLM). We are currently evaluating two models to enhance our response generation capabilities:

- **Model A**: An LLM-based semantic search model.
- **Model B**: An LLM-based semantic search model enriched with knowledge graph facts.

After reviewing 20 queries and their responses from both models, you've seen the potential outcomes of each approach.

The following questions are designed to gather your feedback on the overall performance of the LLM platform and the effectiveness of each model. Please use your observations and experiences from the evaluations you've conducted to answer these questions.

Question 1: How do you feel the platform can contribute to advancing sustainable development in the energy sector? Can you provide an example or outcome?

The direct links to documents and datasets clearly provides a richer source of information to enhance answers from an LLM. As LLM answers from ChatGPT or other tools often resort to generalities when answering questions, it was clear that some of the enhanced answers were using more accurate and contextually informed sets of information to generate a more nuanced answer.

Question 2: Would you use the platform to assist you in daily tasks? Give a reason for your answer.

Yes I certainly would, especially to quickly access information about specific countries, trends, datapoints, or strategies. There is value both in the analytical capabilities of a tool like this to source information from several sources and create a synthesis of the information, as well as the capabilities to act as a direct search engine on a curated set of documents and datasets designed for this specific purpose.

Question 3: Were the responses of the platform up to your expectations? Briefly explain.

Some of the answers were certainly impressive, but there are further ways that the results should be refined. There is often too much focus on UNDP even when this organization isn't mentioned directly in the question. Similarly, even when a question isn't focused on energy, there is an extremely strong focus on energy in the answer. Sometimes the provided information from the documents is focused on in too much detail, eclipsing the more general context so that the analysis isn't as comprehensive.

Question 4: Do you trust the responses of the platform as much as your manual analysis? Briefly explain your answer.

The answers seem to be accurate and well informed. The responses are more reliable when they are directly pulling from proven documents rather than generated purely using an LLM, especially when asking more nuanced or focused questions about specific countries or themes.

Question 5: How would you rate the accuracy of **Model A** and **Model B** in handling domain-specific queries? Were there noticeable differences in factual correctness of these models?

Both answers seemed to be accurate in their responses. One always seemed to stay more general, whereas the other provided more specific details and datapoints. At times the general answer is preferred when asking high-level review questions, and sometimes the specific answer is necessary to provide more detailed data and priorities.

Question 6: In your evaluation, which model demonstrated a better understanding of the context within the domain-specific data? Give a reason.

Both models seem good with context, as the LLM-only has a good general understanding of the content. The country-specific contexts and organization-specific context is definitely enhanced when pulling directly from documents.

Question 7: In your opinion which model (**Model A** or **Model B** or both) can be more helpful for informed decision making in field of sustainable energy? Briefly explain.

The specific details are much more valuable when utilizing a tool like this for professional purposes, since anyone working in the energy field already has a good sense for the general context, and is instead seeking specific datapoints and information contained within the most recent publications and priorities.

Question 8: Did you notice any qualitative differences in the way each model generated responses for queries?

The document focused model jumped quicker into references, but sometimes foregoes the general background information before diving into details. One model seemed to generate paragraphs of information, whereas one had more divisions in the sections. One model seemed to be stumped more often without an ability to answer detailed questions.

Question 9: In your opinion, what are the main challenges and concerns regarding the use of Moonshot AI platform in sustainable energy?

The risks are always that an LLM will take the information provided and try to extrapolate it into a complete picture. As questions become more detailed, this requires the ability for the model to provide citations and direct quotes from trusted information, so that there is never any incorrect or misleading information provided.

Question 10: Are there any ethical or social concerns associated with using the platform in the context of sustainable development? How is the platform addressing these concerns?

If any incorrect information is used for decision-making, this has fundamental concerns. Thus citeability and transparency are key. Also, the tool should never provide specific recommendations or advice, but rather just provide accurate information and insights that could be used when making decisions or policies.

Question 11: Based on your experience, what improvements or enhancements would you recommend for the platform's performance and contribution to sustainable development?

For using the database of documents and datasets, it would be essential to give a user an idea of the extent of the included information, so they know how to query it. Suggestions for queries and interactions with the tool will be essential to enhance the user experience.