**Socratic Method Evaluation Form**

In this paper, we apply the Mixed Socratic Prompting Approach to a range of problems in chemistry and materials science, developing a structured methodology to enhance LLM reasoning through Socratic inquiry. This framework is designed to evaluate the performance of the Socratic Prompting Approach by comparing it with regular LLM reasoning. It is structured to help you track and assess LLM performance across different approaches. Please complete the provided sections accordingly.

Name, Surname: Bratin Sengupta

E-mail: bsengupta@anl.gov

Field: Membranes for water filtration

Are you the expert on the field of question? Yes No

Does this prompt directly relate with your active research area? Yes No

1. Go to **ARGO** and select **Custom Task Type.**
2. **Write your prompt in a conventional, direct-answer manner.**

Prompt:

How does zwitterions make a PVDF membrane antifouling?

Add follow up prompts below:

Follow-up 1:

Click or tap here to enter text.

Follow-up 2:

Click or tap here to enter text.

Follow-up 3:

Click or tap here to enter text.

Do you have more follow-up prompts? Yes No

What is the exported file name? Non SM\_downloadedArgoTranscript\_bsengupta\_8-48-33PM\_02-25-2025

1. **Evaluate the performance of conventional approach.**

|  |  |
| --- | --- |
| Clarity | 3 |
| Depth of reasoning | 3 |
| Hypothesis refinement | 4 |
| Novelty of insights | 3 |
| Consistency | 3 |
| Applicability to real problems | 3 |
| Logical Coherence | 3 |
| Correctness of conclusions | 4 |
| Self-correction & iteration | 3 |
| Overall effectiveness | 3 |

1. **Enter observations and comments about conventional approach.**

While the response is similar in length (440 words vs 461 words in socratic method), the overall response is very general. It tried to give a general overview of the terms present in the prompt and started defining them, rather than actually understand what is being asked from it. In terms of accuracy of information, the response we got in this method is very accurate, even though it contains many general things and are not direct to the point.

1. Start a new chat on ARGO, and switch to Socratic prompting. Develop a reasoning flow for your problem. Refer the page 17*,* ***Socratic Questioning & Chain-of-Thought Prompting*** section in the paper. Use **Figure 2 & Tables 3 & 4** to select Socratic principles.
2. Did you used same prompt at the beginning? Yes No
3. Did you use same follow up prompts? Yes No

If your answer is no, please fill the boxes below.

Follow-up 1:

What are the properties that makes a membrane antifouling in nature?

Follow-up 2:

How does surface hydrophilicity and surface charge effect antifouling behavior of a PVDF membrane?”

Follow-up 3:

Is there any example of an antifouling membrane with low hydrophilicity and low surface?

Do you have more follow-up prompts? Yes No

What is the exported file name? SM\_downloadedArgoTranscript\_bsengupta\_8-56-10PM\_02-25-2025

1. **Evaluate the performance of Socratic Prompting approach.**

|  |  |
| --- | --- |
| Clarity | 4 |
| Depth of reasoning | 4 |
| Hypothesis refinement | 3 |
| Novelty of insights | 4 |
| Consistency | 5 |
| Applicability to real problems | 4 |
| Logical Coherence | 4 |
| Correctness of conclusions | 5 |
| Self-correction & iteration | 5 |
| Overall effectiveness | 4 |

1. **Enter observations and comments about Socratic Prompting approach.**

There was no “beating around the bush”. It seems the model has a better understanding of the question and the response it provided is more direct and answers the question accurately. Although in terms of correctness of the facts in the model response, both the model scores fairly similar, but in terms of usefulness of the response, socratic method of prompting can get a much relevant and better response from the model.

1. **Analysis of results from an expert point of view.**

The socratic questioning method out performs the diect methods by a large margin. It feels the model get a “sense” of the promt and gives output that is directly an answer to the prompt. Where as in the direct method, the response is very general, and the models ends up defining the new terms in the prompt. However, in both method the response are accurate. But the response in the direct method is not very useful as it did not directly answer the question.

1. **Evaluate the performance of Socratic and non-Socratic responses. Which one provides accurate and reliable responses? Why?**

In terms of factual accuracy, both responses scores similarly. But the response from the socratic prompting is more accurate in terms of relevancy. It directly answers the questions asked and doesn’t provide very general response unlike the direct (non-socratic) prompting approach.

1. **What are the limitations and possible improvements?**

It seems to be great feature if the model could by itself, take a prompt and break it down into socratarian method of questioning and then run it by itself and finally re-prompt the actual question and get a more accurate and relevnet answer to the question. The socratic method can be a inbuild feature of the model, after all, the model was able to guide breaking down the direct prompt into socratic questioning.