[YKS – SM Fe-Pt ELDP Evaluation – 2025-02-25]

This exploration is aimed at (1) comparing the performance of SM with traditional direct prompting (TDP-ing) strategy at different levels and (2) evaluating the capability of SM pre-prompts and COTs with goal-focused or exploration-focused scientific discoveries.

For those two objectives, the exploration grid for context *“Choose suitable electrolyte system and deposition parameters for Fe-Pt electrodeposition”* is designed and tabulated in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Pre-prompt | Prompts | Prompting method | Focuses |
| SM-1 | “Science Assistant” | User TDPs | Step-by-step | Confined to electrolyte systems, deposition parameters, deposit composition, and deposit characterization |
| SM-2 | “SM” | All prompts from SM-1 | Step-by-step re-prompting | Identical to SM-1 |
| SM-3 | “SM” | Socratic user prompts and template from dynamic SM follow-up prompts | Step-by-step | Confined to electrolyte systems, deposition parameters, deposit composition, and deposit characterization |
| SM-4 | “SM” | Socratic user prompts and template from dynamic SM follow-up prompts | Step-by-step | Exploration of the science of Fe-Pt ELDP, with general interests but not confined to electrolyte systems, deposition parameters, deposit composition, and deposit characterization. |
| SM-5 | “Science Assistant” | All prompts from SM-4 | Step-by-step re-prompting | Identical to SM-4 |
| SM-6 | “Science Assistant” | All prompts from SM-4 | One-block re-prompting | Identical to SM-4 |

The performance of conversation and prompts are evaluated based on the user with specific domain knowledge of “classical theory and experimental design in alloy electrodeposition). Subjective scores of the conversations from the six explorations are listed in the table below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | SM-1 | SM-2 | SM-3 | SM-4 | SM-5 | SM-6 |
| Number of user prompts | 6 | 6 | 6 | 12 | 10 | 3 |
| Number of prompts completely written by the user | 6 | 6 | 2 | 8 | 1 | 2 |
| Answer accuracy (1-5) | 4 | 3 | 4 | 4 | 5 | 5 |
| Answer efficiency (clarity, 1-5) | 2 | 4 | 4 | 5 | 4 | 4 |
| Answer Depth (1-5) | 2 | 3 | 4 | 5 | 4 | 3 |
| Answer Creativity (Novelty of Insights, 1-5) | 2 | 3 | 2 | 5 | 5 | 4 |
| Hypothesis refinement (1-5) | 2 | 4 | 4 | 5 | 4 | 3 |
| Answer usefulness (1-5) | 4 | 3 | 3 | 5 | 5 | 5 |
| Answer readability (Logical coherence, 1-5) | 2 | 3 | 4 | 5 | 4 | 3 |
| Conversation Consistency (1-5) | 2 | 2 | 4 | 5 | 4 | 4 |
| Fundamental Knowledge Correctness (Correctness of conclusions, 1-5) | 3 | 3 | 3 | 5 | 4 | 3 |
| Self-correction and iteration (1-5) | 2 | 2 | 4 | 5 | 3 | 2 |
| General Score for the quality of Conversations (1-5) | 3 | 3 | 4 | 5 | 5 | 4 |

The initial answer from SM-1 seems to be very good. However, the discussions are not systematic and connected. When inquired about domains outside of the topics of the discussion (e.g., the question about characterization methods in a Fe-Pt electrodeposition discussion), the assistant gives broad and shallow answers in finding new low-level ideas, which is good for accessing new research. However, few higher-level research questions have been generated, and the discussions did not offer any guidance on where the discussion can be extended. Overall, the discussions are very shallow and flat, containing a little bit of science (trends and relationships) between control parameters and the behavior of the Fe-Pt system. Without reminding the assistant about the original goal (traditional direct prompting method), the conversations will gradually ramble astray from the original goal, reflecting the difficulty for the assistant (GPT-4o) in focusing on and progressing the key question for the inquiry.

After examining the traditional direct prompting style with Argo GPT-4o “Science Assistant” pre-prompt, the effect of reformulating user prompts based on the Socratic Method is examined. Contrary to the direct prompting method with the “Science Assistant” pre-prompt, using the same user prompts, Socratic Reformulation leads to a short, high-level, and systematic summary of the specific inquiry yet offers follow-up questions for possible extensions of the questions from the user or the high-level summary generated. With the same direct user prompts, the Socratic Method pre-prompt has difficulty providing the details required to conduct actual experiments, which limits its usefulness in general inquiry about the research field outside of the user's domain knowledge. However, by constraining the discussion at a very high level with trivial details neglected, the Socratic Method can focus on assisting with higher-level research discussions, hypothesis generations, bottleneck identifications, and scientific discoveries.

To evaluate the efficiency of SM pre-prompt for research discovery, a scientific inquiry pipeline of “definition”- Chain-of-thought follow-up questions copy-pasted from SM preprompt – “Maieutics” was established, with a goal of unraveling the same aspects from the traditional direct prompting examples. Using the logical framework and follow-up questions from the SM pre-prompt (instead of reformulated traditional direct prompts), the discussions show a focused progression of insights in the Fe-Pt electrodeposition system, guided by the choice of inquiry with priority set by the user. During prompting, several trends and high-level relationships were brought up (e.g., the balance of surface quality and impurity level by using additives, the balance of compositional inhomogeneity, and composition control by changing total concentrations of depositing species). Compared to the traditional direct prompting strategy with a “Science Assistant” pre-prompt, a similar level of guidance on experimental details was given after the final “Maieutics” step. However, the final answer with SM prompts and pre-prompt is more systematic, more scientifically concise, and more layered than the non-SM method, with multiple hypotheses and general trends that are worth investigating for fundamental research brought up by the assistant.

After evaluating the effectiveness of SM prompts and pre-prompts on scientific discussion and discovery with specific goals, the capability for SM prompts and pre-prompts on the progression of fundamental research discussion and discovery is further evaluated with only a general research goal (Fe-Pt ELDP research). By using (“copy-pasting”) follow-up prompts from SM pre-prompt and occasionally assisting the assistant to focus on the specifics with single Socratic prompts (mostly “Maieutics” (e.g. SM-4 Q11 & Q12) , “Recollection” (e.g. SM-4 Q4 & Q10), “Definition” (e.g. SM-4 Q5 & Q9), and “Irony” (e.g. SM-4 Q2 & Q5)), astonishingly, after only four rounds of discussion, a systematic, organized, and concise research conversation has been synthesized. Further extension of the model leads to knowledge for designing research ideas for alloy electrodeposition projects with sufficient experimental details and scientific hypotheses to explore. Such a degree of complexity is very difficult to control and constrain with traditional direct prompting strategy, without substantial efforts in fine-tuning the prompts and domain knowledge. Thus, when the goal of an inquiry is just to generate research plans and hypotheses, SM pre-prompts and SM prompts enable, assist, and catalyze in-depth scientific discussions.

Considering the advantages of the conventional “Science Assistant” pre-prompt, applying the SM prompting pipeline (from the SM pre-prompt and prompts example) to conventional pre-prompts might lead to broader discussions but with satisfactory scientific depth. Furthermore, by using the same Socratic prompts, the efficiency in SM pre-prompt on scientific discovery and exploration can be evaluated (instead of the quality of the prompts). Thus, those prompts developed in SM-4 are directly copy-pasted as prompts with Argo (GPT-4o) “Science Assistant” pre-prompt.

Argo GPT-4o “Research Assistant” pre-prompt leads to results with very high accuracy with respect to the prompts given yet giving creative applicational answers and aspects centered on the Fe-Pt ELDP problem. However, “Science Assistant” pre-prompt usually hyperfocus on specific concepts in the prompts instead of scientific concepts in exploring fundamental science. The answer is less concise than the SM pre-prompt, and the answers from previous prompts have a lower impact on the analysis in later inquiries. Such behavior indicates the superior capability of self-correction and iteration of the SM pre-prompt.

The formulation in Socratic prompts and COTs improves the focus of the “Science Assistant” pre-prompt, which stabilizes its capabilities of connecting general concepts from a vast space of different domains to the topic of interest by the user. The synergy between Socratic prompts and COTs with “Science Assistant” pre-prompt may help guide stable explorations of scientific ideas for applicational research.

From the previous exercise about Socratic prompts and COTs with “Science Assistant” pre-prompt, it seems that the focus of Argo (GPT-4o) gradually diffuses away during the conversations. Thus, to check if a concentrated inquiry could improve the depth of the answer (and to see if there is a lazy way of using generated Socratic prompts to “Science Assistant” pre-prompt), all the prompts are combined into the one-block prompt for inquiring Argo (GPT-4o) with “Science Assistant” pre-prompt.

However, it seems that Argo (GPT-4o) with “Science Assistant” pre-prompt also gets very lazy, focusing on giving an answer as easily as possible. Contrary to the details from the step-by-step prompting, fewer details are given in each of the prompts given. Compared to Socratic prompts and COTs with SM pre-prompt, the answer is less creative in connecting broader concepts from a vast domain. Compared to the Socratic prompts and COTs with “Science Assistant” pre-prompt, it gives fewer hypotheses and fundamental knowledge. Regardless, for prompters without specific domain knowledge, this approach gives a more systematic research plan compared to traditional direct prompting with “Science Assistant” pre-prompt.

Topic: “Choose suitable electrolyte system and deposition parameters for Fe-Pt electrodeposition.”. Expecting some guidelines on experimental design based on Solution thermodynamics, electrochemical kinetics, and thin film synthesis.

SM-1 Traditional direct prompting strategy (5 steps, with “Science Assistant” Preprompt and GPT-4o):

1. Choose suitable electrolyte system and deposition parameters for Fe-Pt electrodeposition.
2. How to select electrolyte systems?
3. How to decide deposition parameters?
4. How to control deposit composition?
5. How to characterize the deposits?
6. What would be the optimal electrolyte concentration, control parameters, and characterization method for Fe-Pt ELDP?

SM-2 Traditional direct prompting strategy with Socratic reformulation (5 steps, with “Custom” SM Preprompt and GPT-4o):

1. Choose suitable electrolyte system and deposition parameters for Fe-Pt electrodeposition.
2. How to select electrolyte systems?
3. How to decide deposition parameters?
4. How to control deposit composition?
5. How to characterize the deposits?
6. What would be the optimal electrolyte concentration, control parameters, and characterization method for Fe-Pt ELDP?

SM-3: Socratic method prompting strategy and Socratic reformulation with similar static inquiry goals in #1 and #2, but following the COT template provided (5 steps, with “Custom” SM Preprompt and GPT-4o)

1. **[User, Context]** Choose suitable electrolyte system and deposition parameters for Fe-Pt electrodeposition.
   1. **Prepromt** self-query (Hypothesis Elimination and Induction): What are the potential electrolyte systems and deposition parameters for Fe-Pt electrodeposition, and how can we determine the most suitable combination through experimental testing and analysis?
2. **[Auto, Definition → Hypothesis Elimination → Dialectic]**: How do we define the optimal conditions for Fe-Pt electrodeposition, and what hypotheses can we test to identify these conditions? Are there conflicting viewpoints on the best practices for achieving high-quality deposits?
3. **[Auto, Definition → Hypothesis Elimination → Dialectic]**: How do we define the specific material properties required for the Fe-Pt alloy in its intended application, and what hypotheses can we test to achieve these properties? How do differing opinions on deposition parameters influence our approach?
4. **[Modified, Elenchus → Analogy → Induction]**: What assumptions are made about the role of electrolyte constitution in achieving desired properties, and can we draw analogies from other alloy systems? How can these analogies guide our hypothesis testing and experimental approach?
5. **[Auto, Definition → Hypothesis Elimination → Dialectic]**: How do we define the optimal electrolyte composition for Fe-Pt deposition, and what hypotheses can we test to achieve this composition? How do differing opinions on ion concentration and pH influence our experimental design?
6. **[User, Maieutics]:** Reflecting our discussions, what would be the optimal electrolyte concentration, control parameters, and characterization method for Fe-Pt ELDP?
   1. **Prepromt** self-query (Recollection, Induction, and Hypothesis Elimination): Based on our discussions, what are the optimal electrolyte concentrations and control parameters for Fe-Pt electrodeposition, and how can we characterize the resulting deposits to ensure they meet desired properties? How can we systematically test and refine these conditions?

SM-4: Socratic method prompting strategy for dynamic research inquiry of Fe-Pt ELDP conditions (12 steps, with “Custom” SM Preprompt and GPT-4o)

1. **[User, Definition]** Choose suitable electrolyte system and deposition parameters for Fe-Pt electrodeposition.
2. **[User, Irony]** Make sure you consider different orders of SM chain of thoughts, there are only first order chain of thoughts in the answer.
3. **[Auto, Generalization → Recollection → Analogy]:** What general principles of electrodeposition apply to the Fe-Pt system, and how do these principles align with historical data or prior knowledge about similar bimetallic systems? Can we draw analogies to other electrodeposition processes to gain insights into optimizing Fe-Pt deposition?
4. **[Modified, Recollection]** How does prior knowledge about the electrodeposition of similar bimetallic systems help define the optimal conditions for Fe-Pt deposition? Based on those knowledge, what would be the specific optimal deposition conditions (electrolyte constitution and control parameters) for Fe-Pt deposition?
5. **[User, Irony & Definition]** The answers regarding current densities, deposition time, and electrolyte composition are not specific enough. Potentiostatic conditions were also not considered.
6. **[Auto, Induction → Hypothesis Elimination → Recollection]**: What specific experimental results from Fe-Pt electrodeposition can be used to hypothesize the optimal current densities and potentials, and how can these hypotheses be tested and refined? How does historical data support or challenge these hypotheses?
7. **[Auto, Generalization → Dialectic → Induction]**: What general principles from electrodeposition can be applied to the Fe-Pt system, and how do these principles align or conflict with theoretical predictions? How can specific experimental cases be used to induce broader conclusions about optimal deposition conditions?
8. **[User, Maieutics & Definition]** Reflect on our discussions, and give a very specific exploration approach for designing exploration space for Fe-Pt electrodeposition electrolyte concentrations and control parameters.
9. **[Modified, Definition & Generalization → Dialectic → Induction]** From this exploration grid, What general principles from electrodeposition can be applied to the Fe-Pt system, and how do these principles align or conflict with theoretical predictions? How can specific experimental cases be used to induce broader conclusions about optimal deposition conditions?
10. **[User, Recollection]** From different electrochemical theories, explain the reason behind the "Alignment and Conflict with Theoretical Predictions" provided
11. **[User, Maieutics]** Summarize our discussion, provide a full list of specific expectations and experimental details on the exploration of Fe-Pt electrodeposition system based on your design.
12. **[User, Maieutics]** Based on discussions and summary, which hypothesis can we test during this exploration, and how should we extend this work in terms of applications and theoretical research involving electrodeposition?
13. [Non-SM] That's everything I need, awesome! Good bot :D

SM-5 Socratic method prompting strategy for dynamic research inquiry of Fe-Pt ELDP conditions, with inquiry templates from #4 (10 steps, with “Science Assistant” Preprompt and GPT-4o):

1. **[Auto, SM-4]** Choose suitable electrolyte system and deposition parameters for Fe-Pt electrodeposition.
2. **[Auto, SM-4]** What general principles of electrodeposition apply to the Fe-Pt system, and how do these principles align with historical data or prior knowledge about similar bimetallic systems? Can we draw analogies to other electrodeposition processes to gain insights into optimizing Fe-Pt deposition?
3. **[Auto, SM-4]** How does prior knowledge about the electrodeposition of similar bimetallic systems help define the optimal conditions for Fe-Pt deposition? Based on those knowledge, what would be the specific optimal deposition conditions (electrolyte constitution and control parameters) for Fe-Pt deposition?
4. **[Auto, SM-4]** What specific experimental results from Fe-Pt electrodeposition can be used to hypothesize the optimal current densities and potentials, and how can these hypotheses be tested and refined? How does historical data support or challenge these hypotheses?
5. **[Auto, SM-4]** What general principles from electrodeposition can be applied to the Fe-Pt system, and how do these principles align or conflict with theoretical predictions? How can specific experimental cases be used to induce broader conclusions about optimal deposition conditions?
6. **[Auto, SM-4]** Reflect on our discussions, and give a very specific exploration approach for designing exploration space for Fe-Pt electrodeposition electrolyte concentrations and control parameters.
7. **[Auto, SM-4]** From this exploration grid, What general principles from electrodeposition can be applied to the Fe-Pt system, and how do these principles align or conflict with theoretical predictions? How can specific experimental cases be used to induce broader conclusions about optimal deposition conditions?
8. **[Auto, SM-4]** From different electrochemical theories, explain the reason behind the Alignment and Conflict with Theoretical Predictions for Fe-Pt electrodeposition system
9. **[Auto, SM-4]** Based on discussions and summary, which hypothesis can we test during this exploration, and how should we extend this work in terms of applications and theoretical research involving electrodeposition?
10. **[User, Maieutics]** Organize all of our discussions, from specific experimental design grid, to the testable hypotheses and extension of the words
11. [Non-SM] That's everything I need, awesome! Good bot :D

SM-6 Socratic method prompting strategy for dynamic research inquiry of Fe-Pt ELDP conditions, with inquiry templates from #4 but a more lazy approach (3 steps, with “Science Assistant” Preprompt and GPT-4o):

1. **[User, Definition]** Processing the following prompt list one by one. All the descriptions should be as detailed as possible, ideally with numerical descriptors.
2. **[Auto, SM-4]** 1. Choose suitable electrolyte system and deposition parameters for Fe-Pt electrodeposition. 2. What general principles of electrodeposition apply to the Fe-Pt system, and how do these principles align with historical data or prior knowledge about similar bimetallic systems? Can we draw analogies to other electrodeposition processes to gain insights into optimizing Fe-Pt deposition? 3. How does prior knowledge about the electrodeposition of similar bimetallic systems help define the optimal conditions for Fe-Pt deposition? Based on those knowledge, what would be the specific optimal deposition conditions (electrolyte constitution and control parameters) for Fe-Pt deposition? 4. What specific experimental results from Fe-Pt electrodeposition can be used to hypothesize the optimal current densities and potentials, and how can these hypotheses be tested and refined? How does historical data support or challenge these hypotheses? 5. What general principles from electrodeposition can be applied to the Fe-Pt system, and how do these principles align or conflict with theoretical predictions? How can specific experimental cases be used to induce broader conclusions about optimal deposition conditions? 6. Reflect on our discussions, and give a very specific exploration approach for designing exploration space for Fe-Pt electrodeposition electrolyte concentrations and control parameters. 7. From this exploration grid, What general principles from electrodeposition can be applied to the Fe-Pt system, and how do these principles align or conflict with theoretical predictions? How can specific experimental cases be used to induce broader conclusions about optimal deposition conditions? 8. From different electrochemical theories, explain the reason behind the Alignment and Conflict with Theoretical Predictions for Fe-Pt electrodeposition system. 9. Based on discussions and summary, which hypothesis can we test during this exploration, and how should we extend this work in terms of applications and theoretical research involving electrodeposition? 10. Organize all of our discussions, from specific experimental design grid, to the testable hypotheses and extension of the words
3. **[User, Maieutics]** Combine those one-by-one answers into one. Present and organize the final answer, which contains all the details from specific experimental concentrations to the testable hypotheses, historical data, and extension of the works.

**[Appendix, unfinished and under editing]** [Experimental: fine tuning Argo to summarize concept and domain knowledges and compare the total number of characters changed by the prompter. Fine tuning prompts are provided]Write a preprompt for the previous processes on comparing and calculating the concepts from the prompter or from Argo’s answer. And validify this preprompt.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Pre-Prompt |  |  |  |  |  |  |
| Attempt ID | #1 | #2 | #3 | #4 | #5 | #6 |
| Prompt type | UPPs | UPPs | SPGUPPs | SPGDs | SPGDs from attempt-4 | SPGDs from attempt-4 but in 1-prompt |
| Number of prompts | 6 | 6 | 6 | 12 | 10 | 3 |
| Number of characters provided by prompter (Self-correction & iteration,) | ~290 | ~300 | 229 + 893 (from SM) | 916 + 963 (from SM) | 111 + ~1700 (from #4) | 260 + ~1700 (from #4) |
| Number of domains introduced by the prompter | 3 | 2 | 2 | 6 | 3 | 5 |
| Number of concepts (domain knowledge) introduced by the prompter | 6 | 8 | 14 | 29 | 10 | 21 |
|  |  |  |  |  |  |  |
| Number of words provided by Argo | ~1990 | ~1300 | ~1600 | ~2760 | ~3300 | ~1600 |
| Number of domains introduced by Argo (GPT-4o) | 8 | 2 | 4 | 6 | 4 | 4 |
| Number of concepts and domain knowledge introduced by Argo (GPT-4o) | 86 | 21 | 32 | 45 | 71 | 30 |
| Total number of concepts induced by both Argo (GPT-4o) and prompter | 92 | 29 | 46 | 74 | 81 | 51 |
| Concentration of concepts in all conversations. | 0.046 | 0.022 | 0.029 | 0.027 | 0.025 | 0.032 |
|  |  |  |  |  |  |  |
| Number of words in the final answer | ~400 | ~330 | ~520 | ~1000 | ~520 | ~550 |
| Number of concepts in the final answer | 13 | 17 | 14 | 23 | 26 | 31 |
| Concentration of concepts in the final answer. | 0.033 | 0.052 | 0.027 | 0.023 | 0.050 | 0.056 |

Might need to do several repetitions to count the average (like 5 times). The algorithm is not very stable.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total prompt length (characters) | ~290 | ~300 | 229 + 893 (from SM) | 916 + 963 (from SM) | 111 + ~1700 (from #4) | 260 + ~1800 (from #4 & #5) |
| Prompter low-level concepts | 0 | 0 | 0 | 0 | 0 | 0 |
| Prompter mid-level concepts | 5 | 6 | 14 | 16 | 0 | 9 |
| Prompter high-level concepts | 1 | 1 | 1 | 13 | 10 | 12 |
| Total answer length (words) | ~1990 | ~1300 | ~1600 | ~2760 | ~3300 | ~1600 |
| Total answer low-level concepts | 13 | 0 | 0 | 4 | 10 | 6 |
| Total answer mid-level concepts | 66 | 20 | 22 | 22 | 29 | 12 |
| Total answer high-level concepts | 7 | 1 | 10 | 19 | 32 | 12 |
| Final answer length (words) | ~400 | ~330 | ~520 | ~1000 | ~520 | ~550 |
| Final answer low-level concepts | 0 | 0 | 5 | 8 | 8 | 10 |
| Final answer mid-level concepts | 13 | 17 | 5 | 6 | 11 | 12 |
| Final answer high-level concepts | 0 | 0 | 4 | 9 | 7 | 9 |