The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
ASEE
Paper ID #43223

Board 236: Design for Sustainability: How Mental Models of Social-Ecological Systems Shape Engineering Design Decisions

Dr. Andrew Katz, Virginia Polytechnic Institute and State University

Andrew Katz is an assistant professor in the Department of Engineering Education at Virginia Tech. He leads the Improving Decisions in Engineering Education Agents and Systems (IDEEAS) Lab.

Dr. Marie C. Paretti, Virginia Polytechnic Institute and State University

Marie C. Paretti is a Professor of Engineering Education at Virginia Tech, where she directs the Virginia Tech Engineering Communications Center (VTECC). Her research focuses on communication, collaboration, and identity in engineering.

Dr. Tripp Shealy, Virginia Polytechnic Institute and State University

Tripp Shealy is an Associate Professor in the Civil and Environmental Engineering Department at Virginia Tech. He is also the director of the interdisciplinary Sustainable Land Development Graduate Program. His research is focused on helping improve engineering design. He teaches classes about sustainable engineering design, human behavior and infrastructure systems, and adaptive reuse.

Felicity Bilow, Virginia Polytechnic Institute and State University

Felicity Bilow is a graduate student at Virginia Tech.

Novel Approach Designing Interview Protocols with Generative Large Language Models to Study Mental Models and Engineering Design

Abstract

This paper describes the use of AI to support the initial development of an interview protocol designed to elicit engineering students' mental models of socio-ecological-technological systems (SETs) and how these models influence their design decisions. The protocol was created for a study that addresses the need to prepare engineering students to design sustainable solutions suitable for a world afflicted by climate change. Three frameworks informed the creation of the protocol: (1) mental models theory, (2) theory of planned behavior, and (3) social-ecological-technological systems. Given advances in AI and the complexity of the theoretical frameworks, we were interested in learning whether generative AI could support protocol development. We generated questions using the generative text model: Claude-2. These generated questions were ranked by both Claude-2 and a member of the research team, and the rankings were compared. Through this process, we found that generative models can be used to write initial interview questions, but the quality of the questions is not consistent. Specifically, the questions generated were often relevant to the project, but they were not necessarily useful because of the use of awkward language. Despite this, the generated questions served as a helpful starting point for developing a large set of interview questions that were subsequently filtered and refined by the research team.

Introduction

In 2015, the United Nations (UN) [1] developed 17 Sustainable Development Goals (SDGs) with a set achievement year of 2030. The development of the SDGs set an agenda that seeks to improve the quality of life for individuals globally through societal and technological innovation, including providing clean water and sanitation, sustainable cities and communities, and clean and affordable energy to all [1]. One of the SDGs is focused on climate action and the need for individuals to combat the impacts of climate change [1]. Given the nature of engineering work, engineers are certainly a part of this group of individuals; they will be required to confront climate change and design climate resilient systems, as well as help in "restructure[ing] the energy system and related infrastructure into a system that is sustainable, affordable, reliable, and just" [2, p. 740].

Tackling issues like climate change requires an interdisciplinary understanding. It is not just an environmental challenge; it intersects with issues of equity, justice, and global development. Recent studies indicate that engineering students may not fully grasp the significance of non-technical aspects related to climate change and how these aspects relate to their engineering work [3, 4]. Moreover, engineering students often fail to bridge the gap between the technical aspects of engineering and its social dimensions [5, 6]. Recognizing climate change as a symptom of broader social-ecological-technological systems might help make the connections between technical and social problems stronger.

To better understand engineering students' mental models of social-ecological-technological systems (SETS) and how climate change fits within their mental models, we developed an interview protocol. This paper describes the initial development of this interview protocol using generative AI to test the capacity of this rapidly developing technology in supporting qualitiatve research. The protocol integrates concepts from three theories: 1) theory of planned behavior; 2) mental models; and 3) social-ecological-technological systems. The instrument was designed for fourth-year engineering students because these students are at a critical stage in their education where they are preparing to enter the workforce as engineers. By gaining an understanding of how fourth-year engineering students think about technical systems along with social-ecological systems, the goal is to be better informed on how well engineering education is preparing students to think about and design for climate change, and more broadly, consideration for sustainability during engineering design.

Background

Engineering students tend to overlook the life cycle aspects and systems thinking approach of engineering design, focusing primarily on the final product. In a longitudinal study by Kilgore et al. [7] with U.S. engineering students, the authors found that throughout students' second and fourth years, they exhibited minimal consideration for the life cycle thinking of engineering designs and this deficiency did not increase with years of education. More than three quarters of second-year students and again fourth-year students failed to account for maintenance, disposal, or the complexity of design and construction processes.

Engineering students often switch between strategies when faced with complex sustainability problems. For instance, Lönngren et al. [10] investigated the design approaches of third-year undergraduate engineering students. Students either tried to divide and control, simplify and avoid, integrate and balance, or isolate and succumb. The students who were able to use the "integrate and balance" approach were more likely to recognize the interconnected nature of the problem and the necessity of addressing technical, ecological, and social challenges simultaneously.

Engineering students tend to prioritize economic and/or environmental sustainability over social sustainability in design decisions. Mirza et al. [8] conducted a study with first-year engineering students in the U.S., finding that students emphasized economic and environmental aspects of sustainability over social considerations in their design. Similarly, Kelly-Quattrocchi et al. [9] found that designing for the planet often took precedence over designing for people.

Engineering students also tend to prioritize environmental and economic aspects over social aspects in their understanding of sustainable development [11]. Studies by Azapagic et al. [12] and Nicolaou & Conlon [13] indicate that while engineering students demonstrate knowledge of environmental issues like global warming and deforestation, they often lack understanding regarding environmental policies, legislation, and protection and assessment procedures. Engineering students also tend to exhibit various levels of comprehension regarding sustainability, ranging from uni-structural to multi-structural understanding [14, 15]. For instance, research by Carew & Mitchell [14] and Nicolaou & Conlon [13] suggests that undergraduate engineering students often demonstrate either a uni-structured understanding, focusing solely on one aspect of sustainability, or a multi-structured understanding, where they can describe multiple dimensions of sustainability but struggle to articulate the relationships among them.

Student mental models may influence their career plans and professional motivations. For instance, engineering students associate their professional goals more with environmental sustainability and view the implementation of sustainable development as a professional requirement not a personal one [13, 16]. There remains a need to further understand how engineering students perceive sustainability and climate change and how these complex social-ecological-technological systems (SETS) influence not only their career plans but also their professional practice [2]. By gaining insights into students' perceptions of sustainability and climate change, educators and practitioners can better tailor educational approaches and professional development initiatives to foster a more holistic understanding of SETS and enhance students' readiness to address sustainability challenges in their future careers.

Overall Project Objectives

The research presented in this paper is part of a National Science Foundation (NSF)-funded study investigating fourth-year civil and chemical engineering students' mental models of SETS. More specifically, the overall project aims to address the following objectives: (1) characterize engineers' mental models of social-ecological-technological systems (SETS); (2) understand associations between engineers' mental models and their design decisions; and (3) monitor changes in mental models of SETS as students transition from college to industry. To accomplish these objectives, the research study will take place in three phases. Phase 1 is focused on developing an instrument that measures engineering students' mental models of SETS and how these models relate to their design decisions. Phase 2 is not presented here but will involve surveying a national sample of fourth-year civil and chemical engineering students about their mental models of SETS when designing their senior capstone projects. The final phase will follow students from Phase 2 as they transition into the engineering industry to investigate how their mental models change during the first six months of employment.

This paper is a part of the project's first phase and the interview protocol development. Given the arrival of new generative artificial intelligence (AI) tools, we explore how generative text models can help qualitative researchers draft interview questions. Our goal in this paper is to describe this approach to generate an interview protocol as an instrument for measuring mental models of SETS. In the following section, we outline the conceptual approach used to develop the interview protocol.

Conceptual Framework

This study is conceptualized around three frameworks: social-ecological-technological systems framework (SETS), mental models theory (MMT), and the theory of planned behavior (TPB). The social-ecological-technological systems framework characterizes the interactions between humans, the environment or nature, and technical systems, emphasizing the social, environmental, economic, and technical factors associated with the design process [17, 18]. Mental models theory operationalizes how individuals describe, explain, and predict a system's state, form, function, and purpose [19, 20]. For this study, the system is the connection between technical design and social-ecological systems. An individual's mental model then shapes their design decisions (i.e., a behavior). The theory of planned behavior attempts to explain how an individual's attitudes, norms, and perceptions of control predict their intentions and behaviors [21]. Behaviors (i.e., design decisions) then impact social-ecological-technological systems.

Mental Models Theory (MMT)

Mental models are internal cognitive representations of reality that one creates to interact with the world [19, 20, 22] and can be considered as one's perceptions of the material world [19]. These models are a way for humans to describe, explain, and predict a system's state, form, function, and purpose [19, 20]. Individuals use these models to reason and make decisions [22]. We are using MMT to examine engineering students' understanding of SETS and how this understanding impacts their design decisions.

Social-Ecological-Technological Systems (SETS)

SETS is comprised of three dimensions: social systems, ecological systems, and technological systems. Social dimensions can involve any aspects related to humans during the design process; for instance, cultural, economic, and governance factors are typically included in this dimension [18]. Ecological dimensions can consist of a variety of features and processes related to the environment, such as characteristics of the soil, hydrological processes, aspects of the local vegetation, and the weather and climate of a site location [17, 18]. Technical dimensions can include any factors related to the built environment, such as building materials and resources [17, 18]. These three dimensions interact with one another, creating an integrated system [18].

We conceptualize sustainability around the social-ecological-technological systems framework. This approach is similar to other researchers' approaches in the field of urban policy and ecosystems [17, 18] who have used this framework to examine sustainability issues. SETS affords us the opportunity to understand students' perceptions of each system individually, as well as the interactions among the three systems, and the system as a whole, making it appropriate for a study interested in students' mental models of each individual system, as well as the connections between the three systems.

Theory of Planned Behavior (TPB)

According to TPB, an individual's attitudes, subjective norms, and perceptions of control over their behavior predict their intentions [21]. Attitudes are directed toward the behavior and are one's level of (un)favorability toward the behavior [21]. Subjective norms are one's perceptions of the "social pressure to perform or not to perform the behavior" [21, p. 188]. Finally, perceptions of control over the behavior are related to one's perceptions of the difficulty level to perform the behavior [21]. These factors make up an individual's motivation to perform a behavior (i.e., intention), including how much time and work they are willing to put in [21]. Then, a person's perceived behavioral control and intentions predict their behavior [21]. We are using TPB to characterize engineering students' design decisions (i.e., we are treating design decisions as a behavior).

Methods

Question generation using the generative text model

We used a generative text model to help draft questions for an interview protocol that measures students' mental models of SETS. This approach was chosen because it offers a systematic and structured method for generating a large, diverse, and comprehensive question set that probes different aspects of students' understanding and perceptions of SETS. By leveraging the capabilities of a generative text model, we were able to ensure that the interview protocol covers a wide range of topics related to SETS, including their interconnectedness, dynamics, and implications for sustainability. Additionally, this approach allowed us to incorporate insights from existing literature on SETS and sustainability into the interview questions, ensuring their relevance and validity.

We systematically prompted the model Claude-2 to perform two tasks: generate questions and rank the generated questions. These tasks were performed as separate steps. Claude-2 is a generative text model developed by Anthropic. We used Claude-2 because it is a comparatively high-performing model [23]. It is an example of an autoregressive generative text model that generates text in response to input text. This specific model is also an example of a constitutional model that is designed to follow the 3H principles of harmless, honest, and helpful [24].

For each task we instructed Claude-2 to perform, we needed to draft a clear prompt. For the question generation step, we varied the prompt along multiple dimensions. Starting with the definition of mental models as the internal representations people use to (a) describe, (b) explain, and (c) predict the (i) state, (ii) form, (iii) function, and (iv) purpose of a system. We knew we wanted to generate questions for each combination of those two dimensions (i.e., mental model uses and dimensions). We also systematically varied two other parts of the prompt. One part was the system we were asking about. The options were "common across projects" or "specific to student's project". The project team was uncertain about which path would be most informative and functional in the interviews, so we instructed the model to draft questions for each scenario both directed at "common across projects" or "specific to student's project".

The second part was the temperature of the model. In a generative text model, the temperature is a hyperparameter that controls the randomness of the model's output. A higher temperature will result in more randomness and diversity in the generated text, while a lower temperature will result in more likely but less diverse outputs. We instructed Claude-2 to generate questions at four different temperatures: 0, 0.3, 0.6, and 0.9. This meant that we had 96 different prompts for the question generation task (3 mental model uses x 4 mental model components x 2 system types x 4 temperature values). For each permutation of the prompt, we instructed Claude-2 to generate 6 questions. The stem for this prompt is given in Figure 1 (note that curly brackets denote placeholders that were systematically varied)

Act as if you are an expert qualitative social science researcher.
You specialize in writing interview questions for research studies of someone's mental models.
I need your help.
I am trying to write variations of questions for a project to study civil and chemical engineering students' mentalmodels of social-ecological-technological systems and how those inform the students' design decisions.
As you know from your vast expertise, mental models can be defined as the internal representations of the systems to describe, explain, and predict the state, form, function, and purpose of the system.
As such, there are three main dimensions that I need to consider when writing these questions:
1. Mental model dimension - is the question focusing on the state, form, function, or purpose (or some combination) of the system?
2. Mental model use - is the question about describing, explaining, or predicting the state, form, function, or purpose of the system?
3. System type - is the question about social-ecological-technological systems implicated in the student's design project or about systems we have told the students to imagine (if the latter, we need to specify systems that are relevant to both civil and chemical engineering students)?
Someone has already written many of the questions.
I need your help to generate a final set of {number_of_questions} open-ended questions that adheres to the following combination of those three dimensions:

Mental Model Dimension: {mental_model_dim}
Use: {use}
Type: {system_type}

Based on these guidelines and the background information I have given you, generate a set of {number_of_questions} interview questions that adheres to the specified conditions.
Be sure to enumerate your questions as {num_string}.

Start your response with "My suggestions:" and then list your {number_of_questions} questions.
Stop after generating those questions because I do not need additional narrational from you.

Figure 1. Question Generation Prompt

Question ranking using the generative text model

For the second task, we instructed Claude-2 to rank the questions it had generated. The objective here was to test the extent to which the model could narrow down from the list of six to a shorter list that the research team could find useful. We did this by providing the model with the lists of six questions (one batch at a time) and instructing the model to rank the top three questions. As specified in the prompt (see Figure 2), goodness of questions was determined by clarity and likelihood of eliciting insights into participants' mental models. We also prompted the model to include its justification for the suggested ranking. We used this ranking prompt for each permutation of the question generation prompt, meaning we had 96 different groups of questions for the ranking step.

Act as if you are an expert qualitative social science researcher.
You specialize in writing and critiquing interview questions for research studies of someone's mental models.
I need your help.
I am trying to write variations of questions for a project to study civil and chemical engineering students' mental models of social-ecological-technological systems and how those inform the students' design decisions.
As you know from your vast expertise, mental models can be defined as the internal representations of the systems to describe, explain, and predict the state, form, function, and purpose of the system.
As such, there are three main dimensions that I need to consider when writing these questions:
1. Mental model dimension - is the question focusing on the state, form, function, or purpose (or some combination) of the system?
2. Mental model use - is the question about describing, explaining, or predicting the state, form, function, or purpose of the system?
3. System type - is the question about social-ecological-technological systems implicated in the student's design project or about systems we have told the students to imagine (if the latter, we need to specify systems that are relevant to both civil and chemical engineering students)?
Someone has already written many of the questions.
I need your help to rank which questions are best for each of the variations.
The questions I am sending you are about the {mental_model_dim} of the system and are asking participants {use} the {mental_model_dim} of the systems they are working on.
Here is your task:
I will send you a list of questions in the <questions> tag and you will rank the top three of them of the six with 1 being the best question and 3 being the third best question.
A question is good if it is clear, concise, and will elicit a response that will help us understand the mental models of the students.
You will also provide reasoning for your rankings.
You will do this in two sections. The first section will include your rankings and start with "Rankings:" followed by your rankings on new lines.
The second section will include your reasoning and start with "Reasoning:" followed by your reasoning on new lines.
Start your overall response with "My analysis:" and then your rankings section followed by the reasoning section.
Be sure to include the text of the actual question in your response and your reasoning for your ranking.
Here are the questions:
<questions>{questions}</questions>

Figure 2. Question Rankings Prompt

Question ranking by the research team

With the generated questions and rankings in hand, our research team proceeded to the next step in the process. One member of the research team carefully reviewed the model's responses and ranked them according to perceived clarity and conciseness. This rating was on a five-point scale from 1 (poor question) to 5 (good question). The goal of this exercise was to assess whether the model's rankings aligned with the researcher's ratings of questions. This step would help determine the generative model's effectiveness in writing questions that comprehensively cover the domain of social-ecological-technological systems (SETS).

After rating the responses, the team members then compared the research member's ratings to the model's ratings to identify any discrepancies or areas of consensus. By employing this validation process, we aimed to ensure the quality and reliability of the interview protocol generated with the assistance of the generative text model. This iterative approach not only enhanced the completeness and relevance of the questions but also provided insights into the model's performance in capturing questions relevant to the nuances of students' understanding of SETS.

The findings report the question quality, the effect of system temperature, and the effects of prompt components. We report these measures because they provide insights into the performance and functionality of the generative text model utilized in our research methodology. By assessing question quality, we can determine the reliability and effectiveness of the model in generating relevant and insightful questions for an interview protocol to use when studying a specific topic. Understanding the effect of system temperature allows us to explore how variations in model parameters influence the diversity and quality of the generated questions. Similarly, analyzing the effects of different prompt components helps identify the factors that contribute to the success or limitations of the generative text model in question creation.

Findings

In response to the prompts, the model generated a total of 576 questions (in batches of six at a time). An example of such a response is given in Figure 3. The questions were then ranked by the model in a subsequent prompt, and the top three questions were selected for each permutation of the prompt. Examples of the responses from the model for ranking are given in Figure 4.

My suggestions:

Based on your understanding of the function of the social-ecological-technological systems implicated in your design project, what do you predict will be the impact of your project on those systems over the next years?

What aspects of the function of the social-ecological-technological systems related to your design project do you think your project will change in the future? Why?

Can you predict any unintended consequences, positive or negative, that your project may have on the function of the implicated social-ecological-technological systems over time? Please explain.

What new functions do you predict the social-ecological-technological systems associated with your design project may take on in the future as a result of your project?

Based on your understanding of the function of the systems related to your design, what changes in function do you foresee your project directly or indirectly driving over the next decade?

What functionality changes or enhancements would you predict could emerge in the social-ecological-technological systems connected to your work over the long term due to the implementation of your design project?

Figure 3. Example Questions Generated for Predicting Function with Model Temperature = 0.3

My analysis:
Rankings:
1. Based on your understanding of the function of the social-ecological-technological systems implicated in your design project, what do you predict will be the impact of your project on those systems over the next 10 years?

2. What new functions do you predict the social-ecological-technological systems associated with your design project may take on in the future as a result of your project?

3. Can you predict any unintended consequences, positive or negative, that your project may have on the function of the implicated social-ecological-technological systems over time? Please explain.

Reasoning:
I ranked the first question highest because it is the most direct in asking students to predict the impact of their project on the systems' functions over a specific timeframe. This will elicit useful information about their mental models.

The second question is also good at getting at predictions of new system functions, but is slightly less specific by not defining a timeframe. Still it is concise and clear.

The third question is more open-ended but focuses on the critical issue of unintended consequences related to system function. This could reveal insightful information about limitations in the students' mental models.

The other questions are less focused on predictions of function changes or too wordy. The top 3 stand out as most likely to provide useful data to analyze the students' mental models.

Figure 4. Example of a Response to the Question Ranking Prompt

Figure 5 shows the frequency of human rating scores for the mental model dimensions, and Figure 6 shows the frequency of human rating scores for the mental model uses. The color in each plot corresponds to temperature values. The rows of plots correspond to the model's rankings of the questions (i.e., top choice, second choice, and third choice). For example, "Q Ranking 1" was all the questions that the model ranked highest, "Q Ranking 2" was all the questions that the model ranked second highest, and "Q Ranking 3" was all the questions ranked third highest. Looking at the questions in this way illustrates our main findings. First, the model's rankings were not necessarily aligned with the human's rankings. Second, the temperature did not produce any noticeable trends in question ranking by the generative model nor with human ratings. Third, the function and purpose questions seemed to receive higher human ratings than the form and state questions. Fourth, there was better quality among description and explanation questions than prediction questions based on human ratings. More details about these findings are described below.

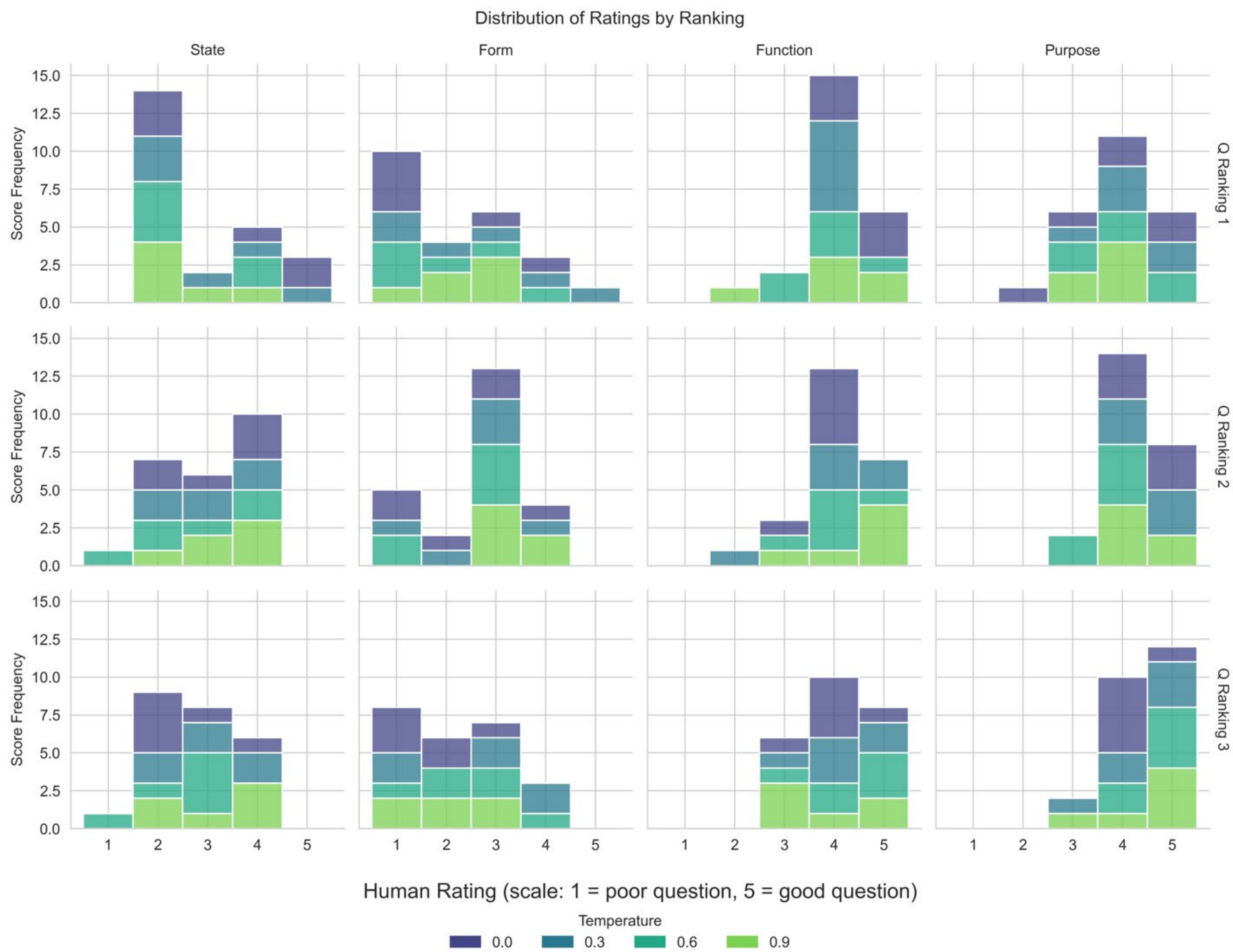

Figure 5. Frequency of Rating Scores vs. by Model Dimensions (State, Form, Function, and Purpose)

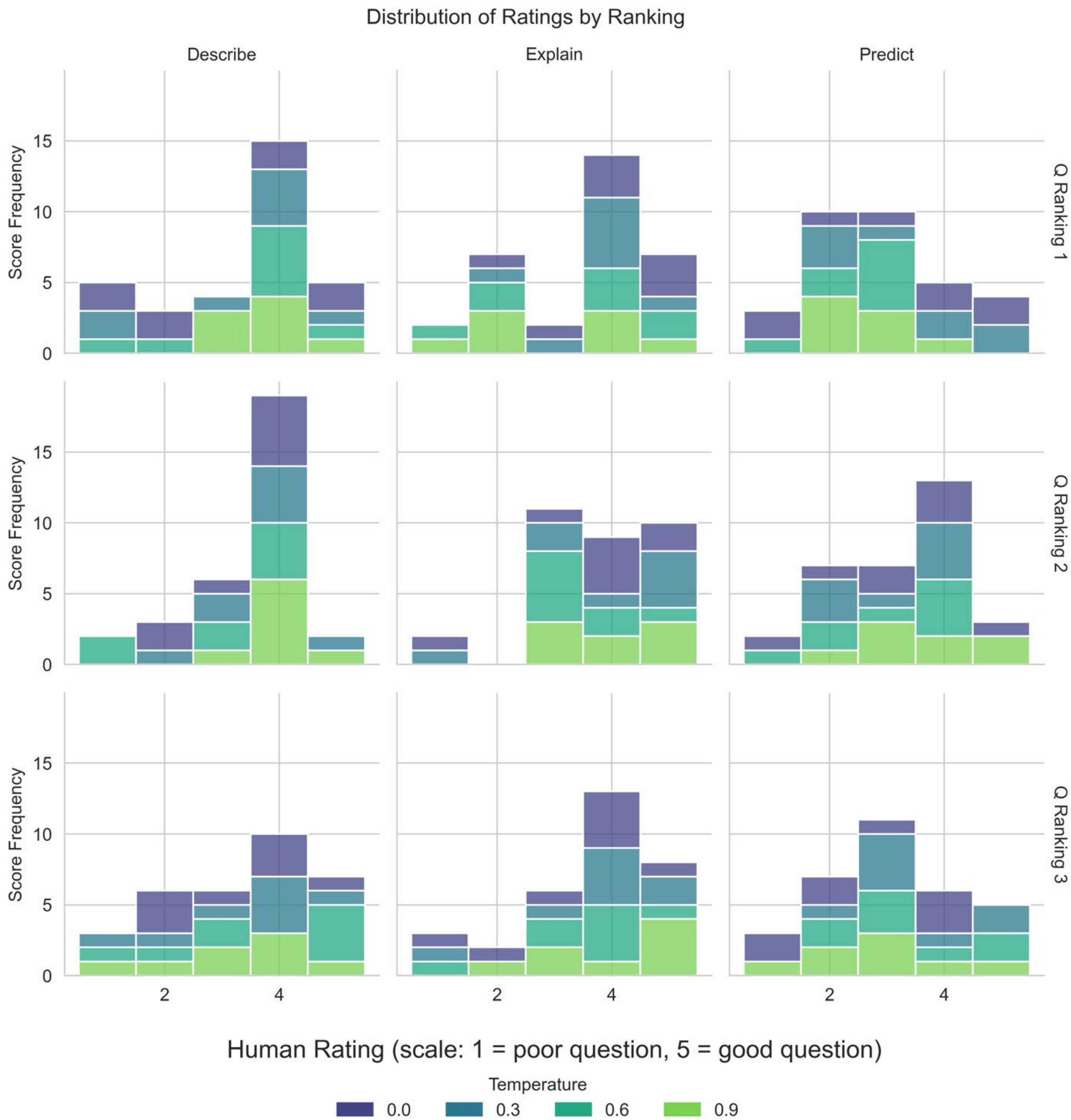

Figure 6. Frequency of Rating Scores vs. by Model Use (Describe, Explain, and Predict)

Question Quality

The model's rankings were not necessarily aligned with the human rater's rankings. This means that whether a question was ranked first, second, or third was not aligned with the human rater's rankings on the five-point scale. However, there was a trend that the rater's top questions were often the second or third ranked question from the model. An instance of this is the question, "Based on your understanding of the social-ecological-technological systems implicated in your design project, what changes or developments do you anticipate in the state or condition of these systems over the next x (select a number) years?" was ranked top by the human rater but third by the model. While the model could generate questions, its alignment with a human reviewer's judgements was not strong.

We also evaluated the quality of the questions generated more qualitatively rather than using the five-point scale. We found that the questions generated by the model were often relevant to the prompt, but they were not always useful for the research team. For example, the model generated questions that were too general or too specific, or that were not relevant to the research objectives. Examples of this included questions that asked about systems in the students' designs rather than systems in which the designs would be embedded. The model also generated questions that were not clear or that were difficult to understand.

Effect of System Temperature

We found that system temperature did not lead to noticeable variation when comparing between the values of 0 and 0.9. For example, "What factors do you think explain the current state of the social, ecological, and technological systems implicated in your design project" was a 0 question, while "Please explain your understanding of the key factors that shape the existing state of the social-ecological-technological system relevant to your design project" was a 0.9 question. Also, there was a notable amount of repetition when moving from one temperature value to the next value, for instance, "What core functions do the systems serve in relation to the problems your design aims to address?" was generated as both 0 and 0.3 questions. Beyond this observation, we did not notice any systematic improvement or decline in the quality of the questions generated by the model as the temperature increased. For example, a 0.6 question was "What unintended purposes or goals do you predict the system you are designing may end up serving, if any?" and a 0.9 question was "If your design was implemented, what purposes or goals do you think it would actually achieve or fail to achieve? Why?" Both questions were rated as good because we thought they were clear, and the interview participant could easily understand them.

Effect of Prompt Components

We found that question quality seemed better when the questions focused on system function and purpose. In addition, we found that the question quality was better among description and explanation questions than prediction questions. Questions asking about state or form were sometimes confusing or irrelevant. For example, a question asking, "In as much detail as possible, please characterize the present state of the social, ecological, and technological systems that relate to the project you are working on." was determined to depend on the student's interpretation of "state" and "relate to the project" and was therefore not useful for the research team.

Discussion

The results of this study suggest that generative text models can be used to generate interview questions, but that the quality of the questions generated by the model is not consistent. Although the questions generated were often relevant to the prompt, they were not always useful for the research team. One potential source of this issue was the awkward language such as "state" or "form" to describe the state or form of the systems. That language may be overly formal and unlike typical vernacular expressed in the model's training data. Based on this possibility, one takeaway lesson is to avoid jargon and favor simple, self-explanatory language.

Another observation was the tendency for the model's second or third-ranked question to be the raters' preferred question of the six options originally generated. This pattern may suggest that the models can mimic some form of preference similar to human raters, but the correlation may not be perfect. Additionally, the reasoning provided by the model for its rankings sounds plausible to human raters. One caveat to this observation is the fact that different human raters might rate the quality of the questions differently. This limitation means that the research team cannot determine how much interrater reliability might affect our findings here.

A third observation was the lack of a systematic improvement or decline in the quality of the questions generated by the model as the temperature increased. This observation suggests that the temperature of the model may not be a significant factor in the quality of the questions generated. However, this observation is limited to the specific model and prompt used in this study, and it is possible that different models or prompts would yield different results. For example, GPT-4 or open-source models will almost certainly exhibit some level of deviation from our findings – whether those are meaningful deviations is an open question. Also, prompts that are simpler, use different language, or direct the model with different instructions also merit future exploration to adopt best practices.

Finally, there were several other limitations for this study not already mentioned. First, we did not test multiple models, so we cannot say if the results are generalizable to other generative text models (e.g., GPT-4). Second, we did not test the model's ability to generate questions for other types of interviews or contexts. This means that we cannot determine if the results would be consistent across different study contexts or whether the combination of asking about mental models and SETSs complicated the model's ability to generate questions. Third, we did not test the effects of different prompts on the model's ability to generate *or* rank questions. Further prompt engineering might yield better results [25]. As such, we consider this study to be an initial foray into the use of generative text models for brainstorming interview questions, and we hope that future research will build on our findings.

Conclusion

In this study, we explored how a generative text model can help draft questions for an interview protocol in a study of student mental models of SETS. We found that generative models can be used to write initial interview questions, but the quality of the questions is not consistent. Although the questions generated by the model were often relevant to the prompt, they were not always useful for the research team due to awkward language. We also found that system temperature did not lead to noticeable variation when comparing among the values of temperature that were tested. Additionally, we found that question quality seemed better when the questions focused on system function and purpose, as well as description and explanation. Despite these limitations, we did find the questions to be a helpful starting point and a way to overcome the cold start problem. We expect that future research can build on our findings and address the limitations of this study by investigating different prompt variations and different study contexts for which to write questions.

References

[1] United Nations, "The 17 goals - Sustainable development." Accessed: Nov. 16, 2023. [Online]. Available: https://sdgs.un.org/goals

[2] M. J. Martin *et al.*, "The climate is changing. Engineering education needs to change as well," *Journal of Engineering Education*, vol. 111, no. 4, pp. 740–746, 2022, doi: 10.1002/jee.20485.

[3] E. A. Cech, "Culture of disengagement in engineering education?," *Science, Technology, & Human Values*, vol. 39, no. 1, pp. 42–72, Jan. 2014, doi: 10.1177/0162243913504305.

[4] A. R. Bielefeldt and N. E. Canney, "Changes in the Social Responsibility Attitudes of Engineering Students Over Time," *Sci Eng Ethics*, vol. 22, no. 5, pp. 1535–1551, Oct. 2016, doi: 10.1007/s11948-015-9706-5.

[5] J. A. Leydens and J. C. Lucena, *Engineering justice: Transforming engineering education and practice*. John Wiley & Sons, 2017.

[6] J. Lönngren, T. Adawi, and M. Svanström, "Scaffolding strategies in a rubric-based intervention to promote engineering students' ability to address wicked problems," *European Journal of Engineering Education*, vol. 44, no. 1–2, pp. 196–221, Mar. 2019, doi: 10.1080/03043797.2017.1404010.

[7] D. Kilgore, A. Jocuns, K. Yasuhara, and C. J. Atman, "From beginning to end: How engineering students think and talk about sustainability across the life cycle," *International Journal of Engineering Education*, vol. 26, no. 2, pp. 305–313, 2010.

[8] T. Mirza, J. Dennison, N. Fila, and S. Purzer, "The sustainability decisions of first-year engineering students," presented at the IL/IN Sectional Conference, American Society for Engineering Education, 2012.

[9] S. Kelly-Quattrocchi, D. Wilson, R. Roberts, and R. Yonemura, "Is protecting the environment all there is to sustainability?," presented at the 2016 ASEE Annual Conference & Exposition, Jun. 2016. Accessed: Dec. 12, 2023. [Online]. Available: https://peer.asee.org/is-protecting-the-environment-all-there-is-to-sustainability

[10] J. Lönngren, Å. Ingerman, and M. Svanström, "Avoid, control, succumb, or balance: Engineering students' approaches to a wicked sustainability problem," *Res Sci Educ*, vol. 47, no. 4, pp. 805–831, Aug. 2017, doi: 10.1007/s11165-016-9529-7.

[11] S. Flament and K. Kövesi, "What do our students know about the future challenges of sustainability?," presented at the SEFI 2020 Annual Conference, Sep. 2020. Accessed: Oct. 17, 2023. [Online]. Available: https://ensta-bretagne.hal.science/hal-03020263

[12] A. Azapagic, S. Perdan, and D. Shallcross, "How much do engineering students know about sustainable development? The findings of an international survey and possible implications for the engineering curriculum," *European Journal of Engineering Education*, vol. 30, no. 1, pp. 1–19, Mar. 2005, doi: 10.1080/03043790512331313804.

[13] I. Nicolaou and E. Conlon, "What do final year engineering students know about sustainable development?," *European Journal of Engineering Education*, vol. 37, no. 3, pp. 267–277, Jun. 2012, doi: 10.1080/03043797.2012.681863.

[14] A. L. Carew and C. A. Mitchell, "Characterizing undergraduate engineering students' understanding of sustainability," *European Journal of Engineering Education*, vol. 27, no. 4, pp. 349–361, Dec. 2002, doi: 10.1080/03043790210166657.

[15] M. K. Watson, J. Pelkey, M. O. Rodgers, and C. R. Noyes, "Exploring student sustainability knowledge using the structure of observed learning outcomes (SOLO)

taxonomy," presented at the 2014 ASEE Annual Conference & Exposition, Jun. 2014, p. 24.583.1-24.583.18. Accessed: Oct. 17, 2023. [Online]. Available: https://peer.asee.org/exploring-student-sustainability-knowledge-using-the-structure-of-observed-learning-outcomes-solo-taxonomy

[16] J. France, J. Milovanovic, T. Shealy, and A. Godwin, "Engineering students' agency beliefs and career goals to engage in sustainable development: Differences between first-year students and seniors," *International Journal of Sustainability in Higher Education*, vol. 23, no. 7, pp. 1580–1603, Jan. 2022, doi: 10.1108/IJSHE-04-2021-0161.

[17] R. P. Bixler *et al.*, "An Observatory Framework for Metropolitan Change: Understanding Urban Social–Ecological–Technical Systems in Texas and Beyond," *Sustainability*, vol. 11, no. 13, Art. no. 13, Jan. 2019, doi: 10.3390/su11133611.

[18] T. McPhearson *et al.*, "A social-ecological-technological systems framework for urban ecosystem services," *One Earth*, vol. 5, no. 5, pp. 505–518, May 2022, doi: 10.1016/j.oneear.2022.04.007.

[19] P. N. Johnson-Laird, "Mental models and human reasoning," *Proceedings of the National Academy of Sciences*, vol. 107, no. 43, pp. 18243–18250, Oct. 2010, doi: 10.1073/pnas.1012933107.

[20] W. B. Rouse and N. M. Morris, "On looking into the black box: Prospects and limits in the search for mental models," *Psychological Bulletin*, vol. 100, no. 3, pp. 349–363, 1986, doi: 10.1037/0033-2909.100.3.349.

[21] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179–211, Dec. 1991, doi: 10.1016/0749-5978(91)90020-T.

[22] N. A. Jones, H. Ross, T. Lynam, P. Perez, and A. Leitch, "Mental Models: An Interdisciplinary Synthesis of Theory and Methods," *Ecology and Society*, vol. 16, no. 1, 2011, Accessed: Feb. 05, 2024. [Online]. Available: https://www.jstor.org/stable/26268859

[23] L. Zheng *et al.*, "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena." arXiv, Dec. 23, 2023. doi: 10.48550/arXiv.2306.05685.

[24] Y. Bai *et al.*, "Constitutional AI: Harmlessness from AI Feedback." arXiv, Dec. 15, 2022. doi: 10.48550/arXiv.2212.08073.

[25] G. Marvin, N. Hellen, D. Jjingo, and J. Nakatumba-Nabende, "Prompt Engineering in Large Language Models," in *Data Intelligence and Cognitive Informatics*, I. J. Jacob, S. Piramuthu, and P. Falkowski-Gilski, Eds., in Algorithms for Intelligent Systems. Singapore: Springer Nature, 2024, pp. 387–402. doi: 10.1007/978-981-99-7962-2_30.