Exploring AI Bots as Simulators in Human Subject Research: A Novel Approach to Ethical and Efficient Experimentation in Engineering Education Research

Johannes Strobel
Teacher Education
University of Texas at El Paso
El Paso, United States of America
ORCID: 0000-0002-2124-1116

Mara Medina
Biological and Biomedical
University of Texas at El Paso
El Paso, United States of America
mnmedina@miners.utep.edu

Emmanuel Sepulveda Guzman
Teacher Education
University of Texas at El Paso
El Paso, United States of America
ORCID: 0000-0001-9442-811X

Maartje van den Bogaard
Teacher Education
University of Texas at El Paso
El Paso, United States of America
ORCID: 0000-0002-2267-3674

***Abstract*— This research paper aimed to evaluate the effectiveness of AI bots in simulating human subjects for research purposes. In recent years, the use of artificial intelligence (AI) in education has gained significant attention. While most approaches explore possibilities for teaching and learning, the possibilities for human subject research have been under-explored. By utilizing AI bots as a tool for data collection and analysis, researchers can potentially reduce costs, increase efficiency, and improve overall accuracy in their studies. This project focused on replicating a traditional qualitative research study on the topic of "humility in engineering" with real human subjects in a simulated study using several publicly available generative AI tools - ChatGPT, Gemini, DebateAI, MasterDebater - to evaluate the effectiveness of AI bots as simulators of human subject research. The project assessed the "personalities" and response patterns of AI tools leading to ethical implications of using AI bots as research simulators. In addition to evaluating the effectiveness and ethical implications of AI bots as simulators of human subject research, the project also explored the potential benefits and limitations of this approach. Overall, this project represents an important step towards understanding the role of AI bots in research and their potential impact on the engineering education research community. By evaluating the use of AI bots as simulators of human subject research, researchers can gain valuable insights into the benefits, limitations, and ethical considerations of this approach. Ultimately, this project seeks to advance the field of research methodology and contribute to the ongoing dialogue surrounding the use of AI in scientific studies.**

Keywords—AI in education research, methodology, simulation

I. INTRODUCTION

Generative AI has been utilized for a wide range of everyday activities. Increasingly, generative AI tools are used to support academic tasks such as the writing of term papers, generating computer code, identifying, and fixing bugs, and increasing efficiency using natural language prompts. Within the academic and science research enterprise, AI tools have been used to generate abstracts and papers [1] to the point that, in healthcare research, the top ten ChatGPT produced publications have been reviewed [2]. Qualitative research management software contains now integrated AI assistance supporting the complex interpretive analysis of unstructured data [3]. AI-driven tools provide content and text analysis [4,5], and research teams continue to experiment with embodied survey bots (AI-controlled virtual avatars) as virtual research assistants conducting interviews [6], yet the use of chatbots as interview participants has not been explored. While there has been extensive research comparing the quality of human-computer vs human-human interaction in conversations [7], most of the work has been done on dialogical features, flow and style of conversations and how nuanced AI systems mimic human speech and conversation patterns [8]. Virtually no investigation has been focused on the use of generative AI tools as a substitute for human subject research. The overall theoretical premise of this paper is to investigate (a) to what degree generative AI tools are capable to replicate information derived from real human subject research collected with interviews, (b) what possible modifications of an "interview" protocol would lead to similar results comparing a study conducted with generative tools to a study conducted with human subjects. Ultimately, this research aims to be a first step to establish generative AI tools as a valid, efficient, and effective substitute for conducting interview research with human subjects and explore the extent of the possible substitution.

II. THEORETICAL AND METHODOLOGICAL FRAMEWORKS

This study is conceptualized as an evaluation research study [9] as the main task is to evaluate research results obtained by human subject research to results obtained by querying several different generative AI tools. As we are not only comparing results from human subject research to a single tool yet multiple generative AI tools, the team chose multiple comparative case study as a methodological framework [10].

III. CONTEXT OF THE COMPARISON STUDY

As contextual comparison, this study using AI chatbot as simulated human subjects was conducted after a study was conducted on humility in engineering with human subjects with the following research questions: (1) How do undergraduate students in engineering define humility and describe embodied behaviors of humility, (2) to what degree is humility valued in

their respective academic environments and (3) what are advantages and disadvantages of humility in the context of learning. This human subject study was conducted through a survey and follow-up interviews with 25 engineering and science students. The study was approved by human subject protocol and participants were recruited at three universities (one Hispanic-serving R1 institution, a predominantly white R1 institution and a R2 technical university). The team used a semi-structured interview protocol. In the context of the comparison study, we only used the results of a small subset of the interview protocol (see Figure I for the Excerpt).

Interview Protocol (used in this comparison study)

What is, for you, being intellectually humble?

Do you value intellectual humility?

Is intellectual humility valued in your current environment?

Is intellectual humility valued in the profession you are preparing for as an engineer?

When you walk into a college, would I find more humble or more not humble students?

Describe a situation where you observed another student behaving with intellectual humility. What exactly did the student do when they behaved this way?

How did this make you feel?

Are there instances where being humble or showing humility helps you as a student? Or holds you back?

Figure 1: Excerpt of the Interview

The interviews were transcribed verbatim using otter.ai and the transcripts were verified by a human to ensure their accuracy. MAXQDA, a qualitative research management system was utilized to manage the coding and analysis process. Analysis was entirely performed by the researchers without the use of the embedded AI support within MAXQDA. The qualitative analysis followed a phenomenographic approach [11] meaning the team aimed to produce the widest possibly variance in the experience of the phenomenon of humility. This framework consisted of open coding followed by axial coding forming categories and themes [12]. As a member of the team read through the interviews, sections of the text were labeled inductively producing codes. For example, a segment of text was selected where the student described a peer acting with intellectual humbleness, and such a situation was given a label corresponding to the specific case. The same process was carried out with the 25 interviews creating a list of codes. Before moving to the axial coding, the codebook produced was checked for redundancies. Axial coding aimed to create categories among the codes produced through open coding, which consisted of grouping codes. For example, a category of definitions grouped all the definitions students provided to intellectual humility. This grouping was followed by memoing, each memo created explained the commonalities among the codes that were grouped into a category. The whole coding process was followed by interrater reliability checks producing commonalities and divergence [13]. Finally, once the axial coding was completed the team determined the themes present that were relevant to the study and produced nuances.

IV. Results of the Human Subject Study

Through open and axial coding, eight major themes were produced from 25 semi-structured interviews with engineering students. In the following result section, the team shares example sources by citing students using pseudonyms of participants and the location in the interview transcript in parenthesis. Students provided a conceptualization of intellectual humility which yielded categories such as openness to scrutiny, self-awareness, open-mindedness, and anti-definitions. In addition, students produced anti-definitions (see Table I for an overview), which included being proud, feeling superior to others, being haughty and/or arrogant. In majority of the cases, it meant not feeling superior to others by bragging or showing pride.

TABLE I. Students' Antidefinitions of Intellectual Humility

Category	
Anti-Definitions	Not bragging (Angel, Pos. 159); Not attaching yourself to your idea (Nathan, Pos. 109); not thinking like you're better than someone (Wesley, Pos. 81); not trying … to put someone else's opinion or perspective down. (Fatima, Pos. 97); snobby (Andrew, Pos. 145); absence of intellectual pompousness (Ben, Pos. 101); not comparing yourself to anyone (Angel, Pos. 163).

Students' value of humility was inquired about, and participants expressed that they are working towards becoming intellectually humble. In the process of becoming intellectually humble, students mentioned that interacting with peers helps them act with humbleness and that they considered it a gradual process. According to participants, they value intellectual humility (IH) because they dislike arrogance among their peers and because intellectual humbleness provides knowledge.

Along these lines, participants expressed their view of their college's value of IH, according to participants' responses, colleges value IH through their services, organizations, and programs, in research, and through some of the faculty. In the case of the college's lack of value of humility, participants said that the college structure does not value humility. This category included the following, master classes, how students learn to compete among themselves, and that college culture leads to unhumble attitudes. These aspects, according to the participants' perspective, indicated the college's poor value of IH. Among students' responses it was possible to create, to their understanding, a list of characteristics of intellectually humble peers, when we asked participants to describe a situation where they observed another student behaving with intellectual humility, they responded: that peers are open to support others, understand the diversity of thought and learning, are open to new ideas, to change, and to learning. Another identified theme was where IH peers are, participants said that humble students are a 'handful,' and only perceived in small gatherings or groups, which also remarks the importance of interaction between peers. On the other hand, students stated that humility depends directly on areas of knowledge, thus, IH was contextualized, given that IH depends on the context of the program and college culture. Among these themes and categories, advantages of being and disadvantages of not being IH were provided by participants when they were asked if there were instances where being

humble or showing humility helped them as students or held them back, see Table II. In the case of advantages, students provided a more nuanced and detailed production compared to the disadvantages of not being IH. Lastly, in the interviews participants were asked if the profession they were preparing for valued humility. These students stated that they hoped for an intellectually humble field and that the value of intellectual humbleness in the engineering field is seen through the value of teamwork, openness to diversity, a friendly working environment, and acknowledging having a mistake from an ethical perspective. Participants also mentioned that the engineering field focused on outcomes, a trait that stems from humility and that the value of humility branches from a utilitarian perspective.

TABLE II. ADVANTAGES AND DISADVANTAGES OF BEING AND NOT BEING INTELLECTUALLY HUMBLE

Advantages of being IH	Disadvantages of not being IH
Enables academic growth, effective teamwork, asking for help, makes approachable, creates a good student-professor relationship, improves interactions with peers, and provides humble discussions with peers and faculty.	Loss of academic growth, waste of time/resources, pride equals stagnation, hinders meeting new people and asking for help, unproductiveness, being closed to other ideas, and lack of people skills.

V. DESCRIPTIONS OF THE AI SYSTEMS

In this project, we used four (4) different available AI tools falling in the broad class of chatbots which are software applications designed to mimic human interaction through voice or text interaction [14]. We choose ChatGPT, Gemini, AI Chat (powered by DeepAI) and MasterDebater. Our selection followed several principles: (1) Widely available with at least a version that is publicly available in order to allow for replication by the community. (2) AI systems which are using different large language models and are based on a different technical backbone of their system (each of the three general purpose AI systems – ChatGPT, Gemini and Chat AI are using different AI technologies). (3) Inclusion of not only general purpose AI systems yet also including a system designed to mimic different human language output – in this case a system that mimics and trains people to debate (MasterDebater). While the human-computer interaction interfaces are very similar (text-based entry and output), the four systems differ in other fundamental ways (see Table III).

VI. DATA COLLECTION THROUGH AI CHATBOTS

Due to the novelty of the approach and the lack of existing protocols, the team developed a systematic process for collecting data from the four AI chatbots. First, the team tested and refined the process with one of the available chatbot systems (ChatGPT) and then expanded the work to other AI tools.

Phase 1 (Explorative): At first, the team queried ChatGPT and Bard (which was rebranded and relaunched as Gemini in the middle of the team's exploration), to conceptualize our constructs of intellectual humility and ask for specific examples and suggestions. In addition, the team asked the systems what happens and consequences of what happens when someone is lacking in humility and the value of humility.

TABLE III. OVERVIEW OF DIFFERENT AI SYSTEMS

AI System	Basis of LLM	Humans involved in training of models	Learning from end-user	Output format	Purpose
ChatGPT	OpenAI GPT foundation models	Yes; RLHF (reinforced learning from human feedback) and supervised learning	No, system is stateless	Intro, Text with bullet points and summary	General
Gemini (formerly Bard)	Gemini and LaMDA Textcorpus	No	Yes preconditioned	Intro, Text with bullet points and summary	General
AI Chat	NLP, DeepCore	Details not publicly described	Yes	Text with bullet points, summary	General
Master Debater	NLP, DeepCore	Details not publicly described	Yes	Brief prose without bullet points	Debate Training

The purpose of this phase was to assess the feasibility of receiving meaningful answers across systems and the extent to which the AI bots provided an answer and their level of elaboration. This also supplied a preliminary basis of comparison of how systems differ in output, making it clear that additional comparison across multiple AI bots would provide a more comprehensive understanding of the capabilities of chat bots in this context. In Phase II (Syntax and Grammar), the team continued to compare results across the second AI system, Gemini, During this step, the team explored the use of words, ordering of words and descriptors as well as syntax/order of words to glean more understanding of the structure of answers and assess which groupings would yield more fruitful results.

In Phase III (development and adjustment of protocol), the team produced a protocol to adjust the original interview

protocol to include idiosyncratic aspects of querying the AI bots, for example, when querying the system how they as students would react to somebody displaying behavior concurrent with humility, the AI systems provided often the answer that they cannot answer the question as they are large language models and not human actors. The team adjusted the protocol to ask more hypothetical questions such as "how would a student respond.." In this phase, the team also discussed and reflected on the differences between each AI system and in terms of mimicking human subject response and the nuance that comes with the "personality" of each system.

In Phase IV (Searching, deciding, and expanding AI systems), the team discussed if ChatGPT and Gemini were sufficient for the exploration and comparison and continued to explore other AI systems. The team searched for additional tools – general purpose and special purpose tools – and explore their value of being included in our search. Initial google searches and feeding brief parts of the protocol into different systems, the team came across several chatbots which seemed different enough in the results and the display of results than just the lists produced by the initial two. As a result, we added AI chat and MasterDebater which provided more conversational results that mimicked human research protocol. One of the chatbots considered. – interview.io – was not included as the tool was developed to provide support for job interviews for technical jobs and less a general interview tool.

During Phase V (Final revision of protocol), the team made final adjustments to the interview protocol by testing the interview protocol in all final four chatbot systems and modifying the overall interview protocol. Modifications include: Edit original protocol for grammar/rearrange grammar to keep consistent across based on preliminary inputs/outputs and experimentation with systems. After testing groupings of questions, the team decided on how each question will be asked or how many questions will be asked to systems to keep consistent across them. In order to keep the focus consistent with the human protocol, the team decided to add the word "engineer" where it seems pertinent or unclear in the interview protocol.

Finally, in Phase VI (Use of protocol, and collecting results), a research team member copied/pasted the questions into the systems one by one or in consistent groups which would provide enough context for the systems and transfer the results into google sheets, highlighting separately questions and answer. All steps regarding input of questions and reported outputs were repeated for all other cues/edited protocols for the remaining chatbots.

VII. Results / Comparison of Results of AI System

In the following sections, the research team presents the findings from the querying of the AI chatbots in comparison to the human subject research. For each interview question we created a separate table. Each table describes (a) answers as given by the four queried AI chatbots, (b) an aggregate section in which the team synthesized in where the AI chatbots (a) produced the same results, (b) what the chatbots added and (c) what the chatbot not mentioned compared to the human subject study.

TABLE IV. Chatbots answers for Definitions of Intellectual Humility

AI System	**What is, for you, being intellectually humble?**	**AI systems (aggregate):**
ChatGPT	Provided a list. Acknowledging limits, openness to learning, respect for others' perspectives, admitting mistakes and uncertainties, seeking collaboration and feedback, embracing continuous improvement.	**What do AI systems do the same:** Provides lists and definitions, as questions being prompted would queue systems to provide, giving similar answers. **What are they adding:** Adds nuance to the typical response of the interviewed college students, which falls along the lines of "trying not to be arrogant" or "thinking that your viewpoint is the only valid one" in a room of your peers. **What did not mention:** What intellectual humility does NOT look like, and how intellectual humility is displayed in the context of being an engineering student or a working engineer in industry.
Gemini (formerly Bard)	Provided a list. Acknowledging gaps, appreciating other viewpoints, continuous learning, prioritizing accuracy over rightness, openness to feedback.	
AI Chat	Provides a definition that includes several of the aspects mentioned by the previous chat bots. And provides a perspective from AI field.	
Master Debater	Provides a definition that includes several of the aspects mentioned by the previous chat bots.	

TABLE V. CHATBOTS ANSWERS FOR VALUE OF INTELLECTUAL HUMILITY

AI System	Do you value intellectual humility?	AI systems (aggregate):
ChatGPT	Values IH and outputs advantages of being IH.	**What do AI systems do the same:** Provide advantages of being IH to support the importance of humility in engineering contexts. **What are they adding:** Chatbots provide a nuanced context of what is humility value and the benefits it brings to the engineering context, students only focused on the academic context while AI included the engineering field. **What did not mention:** human subjects mention they are working towards being IH and said that peer interactions support that process, something that AI did not mention. Thus, human subjects provide a more personal stance on their response and go beyond the literal intent of the question.
Gemini (formerly Bard)	IH is a crucial aspect of someone's functioning effectively. Lists reasons to support, which are advantages of being IH.	
AI Chat	Values IH and outputs advantages of being IH.	
Master Debater	Values IH and explains why this is important in the context of engineering.	

TABLE VI. CHATBOTS ANSWERS FOR VALUED IN ENVIRONMENT

AI System	Is intellectual humility valued in your current environment, as an engineering student?	AI systems (aggregate):
ChatGPT	Outputs that IH is valued among many academic environments. Provides a list to support the thesis, this could also be seen as advantages. (Effective Problem-Solving, Collaboration and teamwork, innovation and creativity, ethical and social responsibility, lifelong learning).	**What do AI systems do the same?** variety of contexts in which IH is valuable. **What are they adding?** draws from a wider, general context (humans drew from their immediate context); adding - unprompted - AI field as context. **What AI systems did not mention?** Different aspects/examples where IH is not valued; humans provided examples; an overall assessment of an environment.
Gemini (formerly Bard)	Responds on terms of the artificial intelligence field. And lists, mitigating bias, continuous learning, transparency and trust, and Human-AI collaboration.	
AI Chat	Responds on terms of the artificial intelligence field. Adds a broader context of technology and society, lists, fosters collaboration, learning, and responsible decision-making.	
Master Debater	In the engineering context IH is valued when it is balanced with a focus on technical proficiency.	

TABLE VII. CHATBOTS ANSWERS FOR SITUATION

AI System	Are there instances where being humble or showing humility helps you as a student? Or holds you back?	AI systems (aggregate):
ChatGPT	Outputs 2 lists, one with benefits/advantages of being IH and another with the challenges.	**What do AI systems do the same?** Provide lists of benefits and advantages. **What are they adding?** AI systems bring more examples of challenges or disadvantages that IH can produce. AI systems are symmetrically balanced by providing the same amount of information for both cases. **What did they not mention?** Not mentioning the distinction between IH helps in school and doesn't help in job search; AI systems answer very literal (students expand their answers and go to other related topics).
Gemini (formerly Bard)	Outputs 2 lists, one with benefits/advantages of being IH and another with the challenges that being IH can produce.	
AI Chat	Outputs 2 lists, one with benefits/advantages of being IH and another with the challenges that being IH can produce.	
Master Debater	Focuses on the disadvantages of showing humility rather than in the advantages/benefits like the other chat bots.	

TABLE VIII. CHATBOTS ANSWERS FOR VALUED IN ENVIRONMENT

AI System	Is intellectual humility valued in the profession you are preparing for as an engineer?	AI systems (aggregate):
ChatGPT	Outputs that is valued in the profession. Provides a list which could also be seen as advantages. (Complex Problem-solving, innovation, adaptability, effective communication/collaboration, ethical and responsible practice, growth and leadership).	**What do AI systems do the same:** Provide lists of how IH is valued by colleagues and moreover, how possessing IH will allow for personal growth of the engineer. **What are they adding:** Both mention and put an emphasis on how displaying intellectual humility would be advantageous in terms of being in the engineering profession. Chat AI also emphasizes how adopting traits of those who display intellectual humility will help engineers, similar to how students describe it in the classroom setting. No situational context/example provided either. **What did not mention:** The systems did not mention how a lack of IH would impact engineers in the workplace. Students are only preparing for the workplace and have not experienced it therefore they cannot provide an answer.
Gemini (formerly Bard)	Outputs that IH is valued in the profession (states that language model is not prepared for career). Provides a list which could also be seen as advantages. (Problem-solving, adaptability, collaboration, learning from mistakes).	
AI Chat	Outputs that IH is valued in the profession. Provides a list how IH can help engineers (foster collaboration, lifelong learning, improve decision-making, enhance problem-solving skills).	
Master Debater	Explains how IH can be valuable in the engineering field.	

VIII. DISCUSSION ON PATTERNS OF CHATBOTS AS SIMULATED HUMAN SUBJECT PARTICIPANTS

As the results in the tables above show a nuanced picture of the extent of the usefulness of AI chatbots to simulate human subject research and the extent the chatbots provide the same or similar level of research results as interviews with human participants, the team found several general patterns within the answers of the AI chatbots which raises several questions and recommendations for other researchers to modify their query protocol if employing a similar approach.

AI systems provide literal answers and do not branch out to other related or interconnected topics. While human research participants entertained different directions to take questions, sometimes veering to related or far related topics, the AI chatbots took a literal stand on the questions and answers were narrowly and exclusively focused on one interpretation of the provided question. This finding has several implications: Asking AI chatbots to answer interview questions could result in a good testbed for how questions are understood and chatbots could be used to test several variations of interview questions and their potential yield in answers.

TABLE IX. CHATBOTS ANSWERS FOR BEAVIOR IN ENVIRONMENT

AI System	When you walk into a college, would I find more humble or more not humble students?	AI systems (aggregate):
ChatGPT	College has a mix of both humble and not-so-humble students. Provides characteristics for both types of students.	**What do AI systems do the same?** Gives characteristics of being humble and not-so-humble. **What are they adding:** Both human subjects and chatbots state that college campus is a mixed situation between humble and not-so-humble students. Human subjects more than chatbots focused on contextualized humility. **What did not mention:** Human subjects focus their responses on campus culture and context and mention specific situations such as competition among engineering programs. And humble peers are only detected on interactions that occur in small groups/gatherings.
Gemini (formerly Bard)	Does not know for sure if we'll find humble or not-so-humble students. Reasons that colleges are a mix of personalities. Then provides factors affecting humility and characteristics of humble students.	
AI Chat	This depends on individual personalities, thus finds it difficult to make a generalization. It is a mix of humble and not humble students. Provides description of both types of students.	
Master Debater	States engineering students are intellectually curious and values diverse perspectives. However, engineering education does not focus on IH and focus on technical knowledge.	

TABLE X. CHATBOTS ANSWERS FOR SITUATION

AI System	Describe a situation where you observed another student behaving with intellectual humility. What exactly did the student do when they behaved this way? How did this make you feel?	AI systems (aggregate):
ChatGPT	Outputs an experience/example of this type of situation (Alex), which included characteristics of an intellectually humble student.	**What do AI systems do the same?** Both provided specific examples. **What are AI systems adding?** Fictional characters vs. real encounters of the humans; some chat systems introduce male and female fictional characters. **What did AI systems not mention?** Missing a wider range of examples; not referring to, elaborating or exemplifying earlier responses; AI systems repeated characteristics and provided fewer behavioral examples.
Gemini (formerly Bard)	A scenario (Nadia and Alex) in which two students disagree. And lists the behaviors of demonstrating humility in this specific scenario.	
AI Chat	Like ChatGPT takes the role of a college student and provides an experience/example of this type of situation (Sarah), which includes characteristics of an intellectually humble student.	
Master Debater	Does not provide a scenario like the other chat bots but provides the context of the engineering field. (An overemphasis on subjective viewpoints may detract from the precision and objectivity required in technical decision-making processes).	

While the discussion with AI chatbots feels like a conversation, AI systems do not refer back or elaborate on previous answers unless prompted. Human research participants have a higher awareness of the context of the interview and refer to earlier statements or connect their answers to earlier conversation pieces. In our prompting for examples of behaviors of humble students, for example, AI systems were able to provide illustrations yet seem weaker and less rich when it comes to real examples and example behaviors.

Regarding implicit biases and engendered answers, examples provided by the chatbots varied in terms of gender. One system provided male and female voices (calling them by engendered names) in their example while one provided male only, and one female only voices when unprompted. Several of our human research participants reflected on potential biases or assumptions in their respective answers during the interview framing their answers and sharing points of view if coming from a different perspective. The team has not seen similar output by the AI-based chatbots and the team is not sure if the system themselves would recognize and subsequently reflect on their biases in their own outputs in the way human subject research participants would.

When asked about how the engineering profession values humility, the chatbots generated more insights on the engineering/profession context compared to the human subjects. As our study participants were primarily undergraduate engineering students, it does not come to a surprise that the AI system is more elaborate and more detailed in its answers. When our human study participants had experience with engineering practice, their answers tended to be concrete and confined to their personal experience while the AI systems answered rather from a global perspective.

It is unclear to the research team to what degree the AI chatbots may limit themselves for providing more humanized responses because they play the role of a language model. In some cases of us prompting, however, chatbots related their answer to the Artificial Intelligence realm, for example when answering "a humble engineer in the field of artificial intelligence would behave like x" adding the "in the field of artificial intelligence" without prompting. In other instances, AI seemed to overrule their own stance that limits them to act as a human subject. When chatbots were prompted to describe specific scenarios their output resembled a human response. For, example, ChatGPT provided an answer in terms of an experience as a student; Gemini warned that as a large language model it couldn't directly observe the world as a human, when AI chat answered as it could observe the world like a human. Even though, in some instance, the chatbots mentioned that they cannot answer due to their inherent limits as a chatbot, they could still answer like a human would. AI chat mentioned, for example, in some prompts that it couldn't answer as a human yet in particular questions it did not provide a limitation and it went directly to say "I can provide an example based on a situation where I observed a fellow student.."

The chatbots followed output patterns such as "systematically balanced" outputs, for example when prompted to talk about disadvantages one would face in displaying humble behavior, the systems answered with a variety of disadvantages yet ended with a paragraph – at least - mentioning the overall advantages of displaying humble behavior. So despite the display of a nuanced profile of the answers, the chatbots ended with a positive outlook. Providing an additional perspective on the same general observation, the AI chatbots modeled how a positive version of civil discourse looks like.

IX. LIMITATIONS, CONCLUSION AND RECOMMENDATIONS FOR FUTURE RESEARCH AND USE

This paper set out to study the possibilities of using AI chatbots as simulators or substitutes for human subject research and investigate the efficiency and ethical considerations. The team's use of chatbots has found a high level of overlap between findings from our human subject research study and the findings from querying the four AI-driven chatbots. In addition, the team's comparison has not found outlandishly divergent results or results from the AI bots which contradicted or disputed research findings derived from interviews of our human subjects. The comparative human subject study utilized a

phenomenographic research methodology focusing on the divergent understanding of participants' experiences. In the context of this approach, the AI chatbot-driven study provided comparable results within a fraction of the time and resources needed than a study with human subjects considering that human subject research has been proven increasingly time-consuming and costly, as it involves recruiting and compensating participants.

It is, however, too early to postulate to what degree our research approach could be replicated for other research frameworks and other research designs. While AI-driven chatbots can mimic human behavior to a certain extent, there are perceived to still lack the complexity and nuance of human cognition and emotions. AI systems are well-reported to contain a variety of biases [see 15 for an overview]. In addition, the basis of the large language models underlying the AI chatbots are often mono-cultural and lack cultural awareness and diversity [16]. These examples indicate potentially severe limitations of the generalizability of research findings as the AI systems are not fully capturing the essence of human behavior.

Another premise of this research endeavor was to investigate the benefits of this approach in the context of increasing concerns of ethical implications of human subject research particularly regarding the vulnerability of participants [17]. Human subject research often raises ethical concerns, such as informed consent, privacy, and potential harm to participants. By using AI-driven chatbots, researchers could potentially avoid ethical dilemmas, as they the large language model systems are not subject to the same rights and protections as human subjects. Research with AI simulators could help researchers conduct studies more ethically and responsibly particularly in reducing harm to human subjects, while still achieving valuable and comparable results.

The use of chatbots as simulators depends on the quality of prompting, the nuances expressed in the prompting and the quality of the understanding of researchers of the human-AI-computer-interaction. AI chatbots could be prompted to react in specific ways to certain stimuli, allowing researchers to test hypotheses and theories in a more controlled environment. This new approach requires additional training of human researchers interacting with the system in not only how to construct prompts, yet also in understanding the limitations of any form of prompts and the capacity of the system to provide meaningful answers.

An immediate application that this team saw was the use of AI chatbots as mindtools or cognitive partners [18] in the research enterprise: AI chatbots can be used to refine interview protocols as they can be early interview test partners. In addition, the answers of the AI systems can help in the production of deductive constructs which can build the basis for deductive code lists. Finally, the AI chatbots can be utilized for exploratory studies helping researchers to either refine or narrow their own research direction.

In conclusion, this project evaluating the use of AI bots as simulators of human subject research has demonstrated some potential for advancements in research methodology. With limitations, researchers could conduct studies more efficiently, cost-effectively, and ethically.

X. Acknowledgment

The research team thanks the human research participants for taking their time sharing insightful and nuanced thoughts on humility in engineering. The team acknowledges the support of the university/college leadership teams who made it possible to reach students and recruit participants for the study. The research team thanks the teams of AI designers and researchers to provide the underlining technology and the human system interfaces that allowed the querying and interaction with the AI chatbots.

XI. References

[1] Khlaif, Zuheir N., et al. "The potential and concerns of using AI in scientific research: ChatGPT performance evaluation." JMIR Medical Education 9 (2023): e47049.

[2] Sallam, Malik. "Bibliometric top ten healthcare related ChatGPT publications in Scopus, Web of Science, and Google Scholar in the first ChatGPT anniversary." JMIR Preprints 10 (2023).

[3] Loxton, Matthew H. "Using MAXQDA's Word-Based Features and AI Assist: Processing Large Numbers of Documents in Strategic Foresight." The Practice of Qualitative Data Analysis: Research Examples Using MAXQDA, Volume 2 (2024): 17.

[4] Salah, Mohammed, Hussam Al Halbusi, and Fadi Abdelfattah. "May the force of text data analysis be with you: Unleashing the power of generative AI for social psychology research." Computers in Human Behavior: Artificial Humans (2023): 100006.

[5] Markowitz, David M. "Can generative AI infer thinking style from language? Evaluating the utility of AI as a psychological text analysis tool." Behavior Research Methods (2024): 1-12.

[6] Hasler, Béatrice S., Peleg Tuchman, and Doron Friedman. "Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds." Computers in Human Behavior 29, no. 4 (2013): 1608-1616.

[7] Hill, Jennifer, W. Randolph Ford, and Ingrid G. Farreras. "Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations." Computers in human behavior 49 (2015): 245-250.

[8] Fryer, Luke K., Mary Ainley, Andrew Thompson, Aaron Gibson, and Zelinda Sherlock. "Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners." Computers in Human Behavior 75 (2017): 461-468.

[9] Powell, Ronald R. "Evaluation research: An overview." Library trends 55, no. 1 (2006): 102-120.

[10] Bartlett, Lesley, and Frances Vavrus. "Comparative case studies: An innovative approach." Nordic journal of comparative and international education (NJCIE) 1, no. 1 (2017).

[11] Åkerlind, Gerlese S. "A phenomenographic approach to developing academics' understanding of the nature of teaching and learning." Teaching in higher education 13.6 (2008): 633-644.

[12] Kolb, Sharon M. "Grounded theory and the constant comparative method: Valid research strategies for educators." Journal of emerging trends in educational research and policy studies 3.1 (2012): 83-86.

[13] Gisev, Natasa, J. Simon Bell, and Timothy F. Chen. "Interrater agreement and interrater reliability: key concepts, approaches, and applications." Research in Social and Administrative Pharmacy 9.3 (2013): 330-338.

[14] Adamopoulou, Eleni, and Lefteris Moussiades. "An overview of chatbot technology." In IFIP international conference on artificial intelligence applications and innovations, pp. 373-383. Springer, Cham, 2020.

[15] Srinivasan, Ramya, and Ajay Chander. "Biases in AI systems." *Communications of the ACM* 64.8 (2021): 44-49.

[16] Wang, Wenxuan, et al. "Not all countries celebrate thanksgiving: On the cultural dominance in large language models." *arXiv preprint arXiv:2310.12481* (2023).

[17] Gordon, Bruce G. "Vulnerability in research: Basic ethical concepts and general approach to review." Ochsner Journal 20.1 (2020): 34-38

[18] Jonassen, David H. "Computers as mindtools for schools: Engaging critical thinking." Upper Saddle River, NJ: Merill. (2000).