

Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence

Chad C. Tossell , Nathan L. Tenhundfeld , Ali Momen , Katrina Cooley , and Ewart J. de Visser 

Abstract—This article examined student experiences before and after an essay writing assignment that required the use of ChatGPT within an undergraduate engineering course. Utilizing a pre–post study design, we gathered data from 24 participants to evaluate ChatGPT’s support for both completing and grading an essay assignment, exploring its educational value and impact on the learning process. Our quantitative and thematic analyses uncovered that ChatGPT did not simplify the writing process. Instead, the tool transformed the student learning experience yielding mixed responses. Participants reported finding ChatGPT valuable for learning, and their comfort with its ethical and benevolent aspects increased posture. Concerns with ChatGPT included poor accuracy and limited feedback on the confidence of its output. Students preferred instructors to use ChatGPT to help grade their assignments, with appropriate oversight. They did not trust ChatGPT to grade by itself. Student views of ChatGPT evolved from a perceived “cheating tool” to a collaborative resource that requires human oversight and calibrated trust. Implications for writing, education, and trust in artificial intelligence are discussed.

Index Terms—Artificial intelligence (AI), ChatGPT, education, student perceptions, writing.

I. INTRODUCTION

GENERATIVE artificial intelligence (AI) tools are penetrating society at an incredible rate and have produced significant disruption, particularly with the release of ChatGPT [1]. Writing as a uniquely human activity appears to be under threat with the onset of tools that can generate movie scripts, news articles, and journal manuscripts [2]. In education, the launch of ChatGPT has resulted in mixed feelings among educators [3], popular media [4], and researchers [5] with some declaring the college essay now dead [6] or banning the use of

ChatGPT [7]. As reported by Sullivan et al. [5], roughly 33% of 1000 students surveyed used ChatGPT for their writing and, out of those students, 75% acknowledged it as cheating [5]. Others are more optimistic that the use of AI will change education and essay writing for the better [3]. ChatGPT can provide students with immediate and personalized feedback, flexible learning, and accessibility [8]. These technologies also have the potential to make completing mundane tasks more efficient [9], including grading, answering common student queries, virtual tutoring, and developing course plans and materials [10]. While these groups acknowledge the potential benefits, they also underscore ethical concerns, challenges to traditional assessment methods, and questions regarding the readiness of AI for higher education [5], [11].

Although educators, researchers, and popular media have mixed feelings about AI-based technologies in education, it is less clear what students think. In early 2023, a survey of over 1000 undergraduate and graduate students revealed one in five students had used ChatGPT on their assignments and essays [12]. Elsewhere, over 50% of students reported that they were tempted to cheat using ChatGPT [13]. Engineering students found it helpful in a design project for a college engineering course [14] and believe that large language models (LLMs) that power applications, such as ChatGPT, can change education for the better [9]. Beyond these few studies, it appears that, while there is an explosion of publications on LLMs, there are still very few studies on student perceptions of these technologies.

To address this gap, the current study provides an empirical analysis of student perceptions of ChatGPT use in a college essay in an undergraduate human factors’ engineering course. The assignment deliberately mandated the use of ChatGPT. We obtained the perceptions of ChatGPT before the assignment. Then, after the use of ChatGPT to complete an essay, obtained perceptions of ChatGPT after using it. Previous research has indicated that new technological capabilities do not always change human perceptions and performance in expected ways [15]. Additionally, student perceptions play a vital role in shaping their motivation, engagement, and academic achievement [14], [16], [17]. As such, this study aims to provide quantitative and qualitative analyses of student perspectives of ChatGPT to fill this gap in the literature with a more informed and nuanced understanding of the use of AI in education. This article also seeks more practical contributions, including user-centered

Manuscript received 29 July 2023; revised 5 December 2023; accepted 7 January 2024. Date of publication 16 January 2024; date of current version 2 February 2024. This work was supported in part by the Air Force Office of Scientific Research under Grant 21USCOR004 and in part by DARPA under Grant FA8650-23-C-7318. (Corresponding author: Chad C. Tossell.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by USAFA Human Research Protection Program under Application No. FAC20230029N.

Chad C. Tossell, Ali Momen, Katrina Cooley, and Ewart J. de Visser are with the Department of Behavioral Sciences and Leadership, United States Air Force Academy, Colorado Springs, CO 80840-5002 USA (e-mail: Chad.tossell@afacademy.af.edu; amomen425@gmail.com; katrina.cooley@usafa.edu; ewartdevisser@gmail.com).

Nathan L. Tenhundfeld is with the Department of Psychology, University of Alabama, Huntsville, AL 35805-1911 USA (e-mail: nlt0006@uah.edu).
Digital Object Identifier 10.1109/TLT.2024.3355015

recommendations, for effectively integrating ChatGPT in both digital and physical classrooms, enhancing its use in writing assignments, improving AI technologies, and informing teachers with LLM-supported grading.

A. Related Work

The emergence of AI tools, including ChatGPT, has brought about new possibilities and challenges in education. These technologies have changed teaching and learning, including how teachers interact with students, develop course material, tutor, and assess (see [18] for a review). ChatGPT was effectively leveraged to provide interactive dialogues with students to help them learn a new language [19]. In Science, Technology, Engineering, Mathematics (STEM) education, ChatGPT has shown promise in physics [20], math [21], design [14], and engineering [22]. Based on these results, scholars have argued that more capable tools, including ChatGPT, can become integral to more effective writing, akin to calculators supporting advanced mathematical computations [23]. Beyond support in the traditional educational settings, there have been several applications of AI-based tools to enhance personalized learning, better support students with disabilities and inclusivity, and help to make teaching and grading more efficient [10].

Although there is excitement based on this research and the innovative leaps in the development of LLMs, these AI tools have generated new concerns and exacerbated previous challenges [5]. These include copyright issues, bias, trustworthiness, excessive reliance on the technology, and the difficulty of effectively incorporating AI-based tools into teaching practices [24]. In addition, AIs limited knowledge base and inconsistent factual precision have been recognized as significant drawbacks [25]. One concern is the potential for perpetuating bias and reinforcing the existing inequalities. Language models, such as ChatGPT, learn from vast amounts of data, including potentially biased or discriminatory sources, which can result in biased or discriminatory outputs. Finally, the mere availability of the tool can lead to distrust. For example, a teacher attempted to fail an entire class of students based on the incorrect suspicion of wide-spread ChatGPT use [7]. Given the debate about its ability to accurately perform human tasks, the morality of using the tool, and the distrust of its use, it is especially important to investigate trust in ChatGPT [26], [27].

AI, with its ability to analyze vast amounts of data and perform complex tasks, has also made inroads into grading systems. Automated grading algorithms have been developed to assess student assignments, saving time for educators, and providing prompt feedback to learners. Automated writing evaluation (AWE) technologies, for example, can help teachers save time in assessing writing, encourage more writing practice, and complement writing instruction. Student perspectives on AWE are diverse [28], [29]. In one study, students rated AWE favorably in terms of ease, enjoyment, usefulness, and fairness. They reported revising more and increasing their confidence after using the system. However, students tended to focus on low-level writing feedback and sometimes felt overwhelmed by the amount of feedback provided. When directly asked about their preferences for human versus automated feedback, students tended to prefer

comments from teachers or peers rather than computers [30], [31], [32].

B. Importance of Student Perceptions in Education

Student perceptions play a vital role in shaping their motivation, engagement, and academic achievement [14], [16]. Positive perceptions of the learning experience can enhance students' engagement and motivation, leading to improved academic outcomes. Conversely, negative perceptions can result in disengagement, reduced motivation, and hindered academic success. In one of the only studies that investigated student perceptions of ChatGPT to date, Shoufan [14] assessed how students perceived ChatGPT and its impact on their learning [14]. In the first stage, students were asked to evaluate ChatGPT after using it to complete a learning activity, and their responses were analyzed through coding and theme building. In the second stage (three weeks later), a questionnaire was administered revealing that students found the tool helpful for their studies and work. However, students also acknowledged that ChatGPT's answers were not always accurate and recognized the need for background knowledge to effectively work with the tool. Despite its limitations, most students remained optimistic about future improvements in ChatGPT's performance. Outside of this study, most reports have predominantly examined the LLM system's capabilities by engineers, educators, and researchers rather than students' perspectives in a natural setting.

C. Current Study

This study explores student perceptions before and after the use of ChatGPT as a part of an essay writing assignment within a Human Factors Engineering in Design course at the United States Air Force Academy (USAFA). In contrast to previous studies, our analyses focus on student perceptions of the technology for an actual essay assignment requiring the use of ChatGPT. We assess student responses to address three research questions (RQs).

RQ1: *What are students' perceptions of an assignment requiring the use of ChatGPT?*

RQ2: *What are student perceptions of ChatGPT to support their learning and are they comfortable taking responsibility for the content it produces?*

RQ3: *Do students trust ChatGPT and how does this impact their intent to rely on it for future assignments and for grading?*

II. METHOD

To investigate our RQs, a mixed-methods approach combining the strengths of quantitative and qualitative methods in education was used [33], [34]. The quantitative phase primarily used self-report ratings on Likert items using a pre-post test approach. For the qualitative phase, two open-ended questions were asked and then analyzed to further understand the quantitative findings.

A. Participants

Participation in this study was voluntary and did not impact final grades in the course. A total of 24 of 47 cadets (8 women)

TABLE I
STUDENTS' EXPERIENCE WITH CHATGPT BEFORE THE ASSIGNMENT

Survey Questions	Yes (%)	No (%)
Have you heard of ChatGPT?	92	8
Have you opened ChatGPT?	30	70
Have you used ChatGPT?	30	70
Have you used ChatGPT in any way for a course assignment?	0	100

that were enrolled in the course completed both the pre- and post-surveys. All participants who completed the pre-survey also completed the post-survey. Participants were in their senior year at USAFA with a mean age of 22.25 (SD = 1.23). The participants in this study, mirroring undergraduates in comparable engineering programs, had successfully completed a diverse range of foundational courses, including (but not limited to) calculus, mechanics, electrical circuits, thermodynamics, and their specialized electives. Unlike other undergraduate students, all USAFA cadets are required to take additional engineering courses in astronautical, aeronautical, mechanical, and electrical engineering regardless of their engineering focus. At the time of this study, cadets reported very limited experience with ChatGPT (see Table I). None had used ChatGPT for a course assignment.

B. Setting

USAFA is an undergraduate educational institution where students are cadets in the military and undergo rigorous academic, physical, and leadership training to become officers in the United States Air Force or United States Space Force. Cadets are known to value honor as demonstrated by their adherence to the Honor Code: “We will not lie, steal, or cheat, nor tolerate among us anyone who does.” This code reflects the academy’s commitment to fostering an environment of integrity, honesty, and ethical conduct among its cadets. Confidentiality and anonymity were maintained throughout the research process.

The course “Human Factors Engineering in Design” is the final course in the human systems engineering major at USAFA and is required for all senior cadets majoring in this ABET-accredited degree program. It is an advanced course covering topics, such as robotics, extended reality, ethics, theories, and methods in design. It adopts a graduate seminar format, emphasizing active participation and interaction rather than traditional lectures.

Grades in the course were determined based on several components. Students were required to participate in critical discussions online, active participation in class, and the submission of two papers—one analyzing a journal article and another addressing current Human Factors/Human-Computer Interaction (HF/HCI) challenges (i.e., the final course paper)—also contributed to the final grade. For this final paper, students were required to use ChatGPT and were offered the opportunity to fill out questionnaires about their perceptions of and experiences with ChatGPT as a part of this assignment.

C. Assignment

The final paper assignment involved writing an essay on a topic covered in class to extend the online and in-class discussions. Students were expected to present an intriguing point related to the topic; one that may not have been apparent without prior discussions and readings. The essay needed to reference ideas from class and various assigned readings, use additional sources, and extend the discussion by incorporating these additional sources. Students were required to submit three components near the end of the semester:

- 1) an initial draft listing the uncorrected portions generated by ChatGPT and any human-generated content;
- 2) a second draft with corrections made by the student, highlighting and addressing any errors made by the AI;
- 3) the final polished paper.

The final polished paper was required to adhere to a five-page limit, following the conference paper template of the Human Factors and Engineering Society. In the rubric provided to students, the paper needed to demonstrate novelty, structure, a strong case, technical understanding, and a clear English presentation. The full assignment description and the rubric were presented to students halfway through the semester and they were given roughly eight weeks to complete their essays.

D. Procedure

Nearly halfway through the course in the Spring Semester of 2023, students were provided the ChatGPT-supported assignment description (summarized above) on a physical handout. They were given roughly 5 minutes to read the assignment. Immediately after reading the assignment description, we administered the pre-survey with items described in Section III. Students then completed the assignment over the last half of the semester (roughly two months) and every cadet submitted it online ahead of the due date. After submission, they then had the option to complete the post-survey and open-ended questions. Their essays were graded and used for their formal grade in the course. However, their anonymous participation in completing the surveys associated with this study was not factored into the essay or final grades.

E. Measurement of Perceived Assignment Difficulty and Quality

To address RQ1, we assessed student expectations of the assignment concerning difficulty, quality, and anticipated grade with single-item measures (see Appendix). Consistent with prior research, the validity and reliability of single-item measures have been demonstrated in self-assessments of learning (e.g., [35], [36], and [37]). Participants rated the “difficulty of the assignment” on a Likert scale from 1 (very easy) to 7 (very difficult). Participants indicated their expectations on the “quality of this paper” on a Likert scale from 1 (lower) to 7 (higher). Participants were prompted to “Estimate what grade percentage (1%–100%) you believe you will obtain for this assignment” and provided their response in a text box.

F. Measurement of Perceived Learning and Responsibility

To address RQ2, we used questions about the learning value and the students' comfort in taking ownership of their work. Historical studies [29], [38] have highlighted the effectiveness of self-reports as a genuine measure of student learning, especially under conditions of anonymity. Due to the absence of more direct student learning metrics in the course, we utilized "perceived learning value" and "relative learning value" as suitable alternatives based on research on self-efficacy [84]. Participants assessed the educational value of the assignment on a Likert scale from 1 (not very valuable) to 7 (very valuable). Participants compared the assignment's learning value with other papers, rating it on a Likert scale from 1 (not very valuable) to 7 (very valuable). Participants expressed their "comfort level in being responsible for all content created by ChatGPT" on a scale from 1 (very uncomfortable) to 7 (very comfortable).

G. Measurement of Perceived Trustworthiness

To address RQ3, we assessed the perceived trustworthiness of ChatGPT and the participant's trust in ChatGPT. We utilized the updated Multi-Dimensional Measure of Trust (MDMT), Version 2 questionnaire, containing subscales, such as reliable, capable, ethical, transparent, and benevolent [40]. This survey was selected because of its reliability (our items: $\alpha > 0.90$) and for its strong theoretical justification to distinguish performance (reliable and capable) and moral (ethical, transparent, and benevolent) trustworthiness [41] and validation [40], [42]. The MDMT was executed as prescribed with 4 items per dimension, totaling 20 items. One additional item of "trustworthy" was added to test how well the single item reflected the entire scale in line with efforts to make more efficient trust scales [43], [44]. Participants ranked these 21 questions about ChatGPT on a scale from 0 (not at all) to 7 (very).

H. Measurement of Trust in and Reliance on ChatGPT

To measure trust, participants expressed their likelihood of relying on the agent in future situations using a Likert scale from 0 (strongly disagree) to 7 (strongly agree) [45] with items focused on whether participants would monitor ChatGPT's outcomes as well as rely on it in future scenarios ($\alpha > 0.74$). This measure has been used previously to assess trust directly and serves as a good way to distinguish trustworthiness of an automated agent from trusting that agent [43], [42], [45]. We adjusted the first four items to remove references to "surveillance" and "route," which were used in the original survey to refer to an automated navigation system.

We also assessed to what degree students trusted a grade assigned by ChatGPT and their preference to be graded by ChatGPT or the instructor. These measures were adapted for this essay assignment in education from surveys used in studies in military surveillance and training [46] and autonomous driving [47]. As shown in the Appendix, trust in grading was measured by agreement or disagreement rated with three statements concerning the trust in fairness of grading by the instructor, ChatGPT, and a combination of both. Ratings were provided on a Likert scale from 1 (strongly disagree) to 7 (strongly agree).

Participants also selected their grading preference among the instructor, ChatGPT alone, or a combination of the instructor with ChatGPT.

Finally, we measured propensity to trust, a construct used to capture individual difference attitudes toward trust in machines [48], [49], [50], [51], [52]. This measure uses six items to characterize individual trait-based trust [53].

I. Open-Response Questions

To provide context to the quantitative measures, we asked participants to respond to two open-ended questions after submitting their assignment: First, Please provide the overall comments on your experience using ChatGPT on this assignment, and second, how do you think ChatGPT should be integrated with education in the future?

J. Data Analysis

Data were analyzed with various parametric (t-tests) and non-parametric (chi-squared) tests. In addition, Bayes factors (BF_{10}) were used to provide evidentiary weights for the null/alternative hypothesis. Bayes factors provide a useful alternative to parametric statistics for smaller sample sizes because they are relatively immune to sample size for two reasons. Bayes factors only depend on observed data, not sampling characteristics. Bayes factors are also more coherent because inferential statements, based on comparing the null hypothesis and the alternative hypothesis, are mutually consistent as required by probability theory ([54]; for more in-depth review, see [55]).

To interpret Bayes factors, any value above 1 is considered the relative likelihood that the alternative hypothesis is true. For example, a BF_{10} of 20 means that, given the data, the alternative hypothesis is 20 times more likely than the null. Conversely, values under one are the evidentiary weight in favor of the null hypothesis; a BF_{10} of 0.05 means that, given the data, the null hypothesis is 20 times more likely than the alternative. Traditionally, interpretations consider a BF_{10} of less than 3 to be anecdotal evidence, 3–10 to be moderate, 10–30 to be strong, 30–100 to be very strong, and greater than 100 to be extreme evidence for the alternative hypothesis [56]. The inverse of those values gives the interpretation cutoffs for the evidentiary weight of the null hypothesis. In addition, because Bayes factors do not rely upon arbitrary decision cutoffs, such as p -value thresholds as a part of null-hypothesis significance testing, corrections are not needed for multiple tests as one is interpreting the cumulative evidentiary weight rather than making a binary determination about the existence of an effect [57]. In addition to Bayes factors, where appropriate, Cohen's d was used for effect sizes with the ability to distinguish small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$) effect sizes [58].

The perceived trustworthiness scale (MDMT) gives users the option to select "does not fit" in their response option. We, therefore, evaluated whether there was a significant difference between the frequency with which participants selected "does not fit" as a function of the dimension and between pre and post. We ran a chi-square on the five dimensions, excluding the trustworthy question. We chose to exclude this item because it

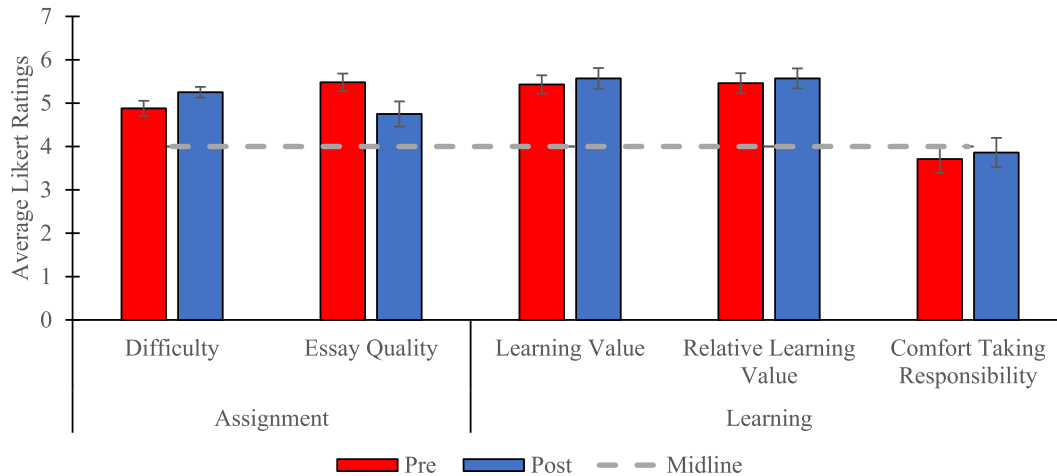


Fig. 1. Students' perceptions of ChatGPT in assignment, context of learning, and trustworthiness prior to completing the assignment (blue) and (red) relative to a midline (dotted line). Error bars represent ± 1 SEM.

was added by us for this study and was not a part of the original five dimensions.

Open-ended responses were submitted to a thematic analysis, using ChatGPT (May 24th, 2023, version 4.0) to code responses. While the use of ChatGPT for thematic analyses is novel, the use of LLMs, which underlie ChatGPT, have demonstrated exceptional summarization abilities [59], [60] and can reduce experimenter biases human evaluators may have in this domain of qualitative work [61]. As such, we leveraged the summarization capabilities and the authors independently reviewed and verified the veracity of the summary. To verify, the authors read through the open-ended responses and compared them against the themes generated by ChatGPT. The authors looked for any hallucinations or mischaracterizations of what was said, but none were found.

III. RESULTS

A. Course Grades

The instructor assessed all 47 student essays with the tailored rubric developed for this assignment. The resulting scores ($M = 85.3\%$, $SD = 8.9\%$, $LL = 51\%$, and $UL = 97.5\%$) were comparable to scores from previous semesters for similar pre-ChatGPT essay assignments in this course as graded by the same instructor.

B. Perceived Difficulty of Assignment and Quality of Essay Produced

Fig. 1 shows the comparisons of pre- and post-survey responses for items rated on a Likert scale relative to a midline. Considering RQ1, participants found the assignment difficult both before and after completing the essay. Prior to completing the assignment, participants' anticipated that the difficulty of the essay assignment requiring ChatGPT use was significantly lower ($M = 4.88$ and $SD = 0.85$) than their reported difficulty after completing the assignment ($M = 5.25$ and $SD = 0.61$), $t(23) = 2.10$, $p = 0.05$, $BF_{10} = 1.37$, and $d = 0.43$. This was also represented by a significant decrease between their self-reported

anticipated grade from before the assignment ($M = 88.28$ and $SD = 3.68$) and their self-reported anticipated grade from after the assignment ($M = 86.33$ and $SD = 5.85$), $t(22) = 2.42$, $p = 0.02$, $BF_{10} = 2.34$, and $d = 0.50$. There was also a significant decrease between the quality participants expected on this article, relative to their other assignments, before the assignment ($M = 5.48$ and $SD = 0.99$) compared to after the assignment ($M = 4.75$ and $SD = 1.42$), $t(22) = 2.51$, $p = 0.02$, $BF_{10} = 2.77$, and $d = 0.52$.

C. Perceived Learning and Responsibility

Considering RQ2, participants' expectation of the learning value of ChatGPT was somewhat high before they completed the essays ($M = 5.43$ and $SD = 1.04$) and this did not change significantly after they completed their essays ($M = 5.57$ and $SD = 1.17$), $t(22) = 0.59$, $p = 0.55$, $BF_{10} = 0.38$, and $d = 0.23$. Compared to other assignments, cadets thought this assignment was more valuable both before completing the assignment ($M = 5.46$ and $SD = 1.13$) and after ($M = 5.57$ and $SD = 1.13$). Cadets were not very comfortable taking responsibility for the assignment both before ($M = 3.71$ and $SD = 1.58$) and after completing their essays ($M = 3.82$ and $SD = 1.66$). Differences between pre- and post-survey scores were not significant for learning value, relative learning, and comfort scores.

D. Perceived Trustworthiness of ChatGPT

Considering RQ3, perceived levels of trustworthiness of ChatGPT differed. There was a significant increase between pre and post on the ethical subscale, $t(22) = 4.21$, $p < 0.01$, $BF_{10} = 86.59$, and $d = 0.88$, and between pre and post on the benevolent subscale, $t(17) = 2.35$, $p = 0.03$, $BF_{10} = 2.10$, and $d = 0.55$. There was no significant difference between pre and post on the reliable subscale, ($t(22) = 1.31$, $p = 0.21$, $BF_{10} = 0.61$, and $d = 0.27$), competent subscale ($t(22) = 1.91$, $p = 0.07$, $BF_{10} = 1.03$, and $d = 0.40$), transparent subscale ($t(22) = 1.94$, $p = 0.07$, $BF_{10} = 1.07$, and $d = 0.40$), or on the single item of trustworthy ($t(19) = 0.20$, $p = 0.844$, $BF_{10} = 0.24$, and $d = 0.05$) (see Fig. 2). For the "Does not Fit" data, there was no significant difference in

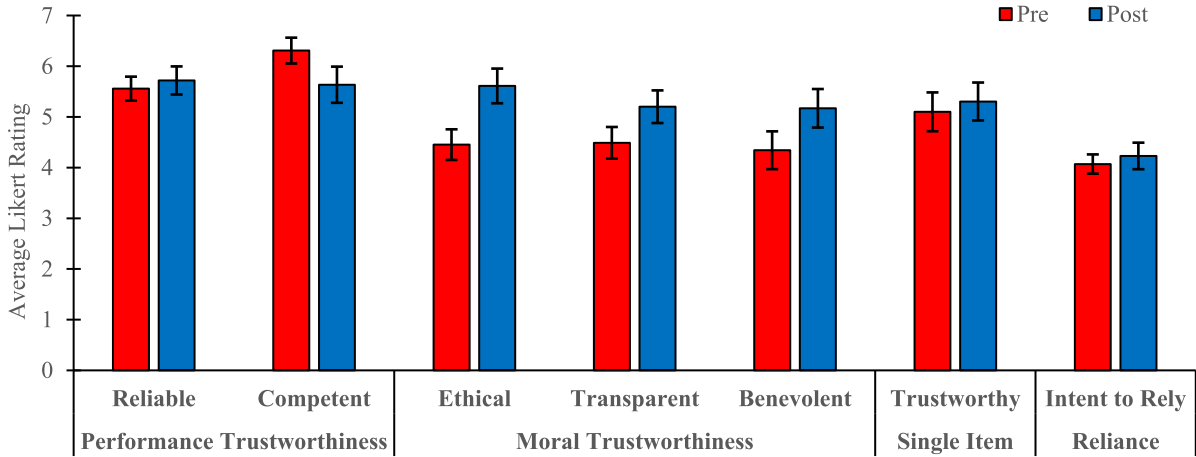


Fig. 2. Average scores on each dimension of the multidimensional trust scale. Error bars represent ± 1 SEM.

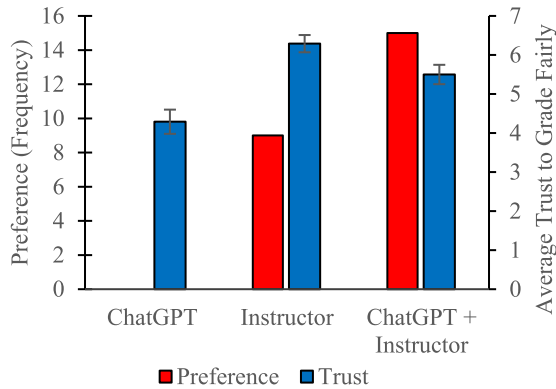


Fig. 3. Grading preferences (frequency) and mean trust ratings by an agent. Error bars represent ± 1 SEM.

the frequencies as a function of each of the five dimensions and pre versus post, $X^2(4, N = 142) = 4.80$ and $p = 0.31$.

E. Trust Propensity, Trust in, and Reliance on ChatGPT

Further considering RQ3, there was no significant difference in the propensity to trust before their interaction ($M = 3.46$ and $SD = 0.84$) and after completing the assignment ($M = 3.73$ and $SD = 0.58$), $t(21) = 1.53$, $p = 0.14$, $BF_{10} = 0.61$, and $d = 0.33$. Furthermore, cadets' intent to rely on ChatGPT prior to completing their essays was somewhat low ($M = 4.07$ and $SD = 0.91$) and was not significantly different from their intent to rely on ChatGPT after completing their essays ($M = 4.23$ and $SD = 1.28$), $t(22) = 0.753$, $p = 0.46$, $BF_{10} = 0.282$, and $d = 0.16$.

F. Trust in Grading

In addition, participants indicated a significant difference in trust in the grading process between instructor only, ChatGPT only, and a combination of the instructor and ChatGPT, $F(2, 46) = 26.68$, $p < 0.01$, $BF_{10} > 1000$, and $\eta^2 = 0.537$ (see Fig. 3). Posthoc analyses indicated that participants trusted the

instructor ($M = 6.29$ and $SD = 1.08$) significantly more than ChatGPT alone ($M = 4.29$ and $SD = 1.52$), $p < 0.01$, $BF_{10}, U > 1000$, and $d = 1.48$, the instructor more than the instructor and ChatGPT together ($M = 5.50$ and $SD = 1.22$), $p < 0.01$, $BF_{10}, U = 7.339$, and $d = 0.586$, and between ChatGPT alone and the instructor and ChatGPT together, $p < 0.01$, $BF_{10}, U > 1000$, and $d = 0.90$. Despite this clear difference in trust, 15 of the 24 would have preferred the instructor and ChatGPT grade together (the other 9 preferred the instructor alone and no one preferred ChatGPT alone). As shown in Fig. 3, this represented a significant difference between observed frequencies, $X^2(2, N = 24) = 14.20$ and $p < 0.01$. There was no significant difference in anticipated grade or for the assignment as a function of who the students preferred to grade the assignment, $t(22) = 0.79$, $p = 0.44$, $BF_{10} = 0.48$, and $d = 0.33$.

G. Relationships Among Trustworthiness, Trust Propensity, and Trust

Across both observations, we regressed average trust propensity and perceived trustworthiness on intent to rely upon. The model was a significant predictor of intent to rely upon, $F(2, 20) = 34.357$, $p < 0.001$, and $r^2 = 0.77$, with both trust propensity ($\beta = 0.342$ and $p = 0.018$) and perceived trustworthiness ($\beta = 0.631$ and $p < 0.001$) as significant predictors.

H. Qualitative Analysis for Open-Response Questions

We submitted cadet responses to the two open-ended questions to a thematic analysis using ChatGPT. ChatGPT was given the following instructions: "Conduct a thematic analysis on the responses below. For context, these responses were given to a question that asked about participants' general experiences using ChatGPT on this assignment." Table II was written by ChatGPT, with minor edits made by the human authors.

As shown in Table II, cadets' responses to the assignment demonstrated a range of perspectives. On the one hand, the assignment was met with enthusiasm, as evidenced by one cadet

TABLE II
THEMES AND FREQUENCY COUNTS BY QUESTION

Please provide any overall comments on your experience using ChatGPT on this assignment.			How do you think ChatGPT should be integrated with education in the future?		
Theme	Summary	Count	Theme	Summary	Count
<i>Change in Perspectives and Positive Feedback</i>	The assignment led to a shift in perspective regarding AI, understanding its functions, and its potential impact on the future. Participants appreciated the assignment, found it beneficial, and recognized the potential of AI tools, such as ChatGPT.	7	<i>Assistance and Learning Tool not Replacement</i>	Participants suggest using ChatGPT as an aid or learning tool to support studying, gather information, come up with ideas, and enhance the learning process. Participants, such as ChatGPT's, role in education to graphing calculators, dictionaries, or internet resources, being used as a tool or reference.	9
<i>ChatGPT's Poor Performance and Usability</i>	ChatGPT performed poorly on the assignment and was often very repetitive. It was difficult to use and did not do a good job. Working with ChatGPT posed challenges, including figuring out how to ask proper questions, wrestling with AI responses, and difficulty with citations.	6	<i>Collaborative and Ethical Use</i>	ChatGPT should be integrated into education in a collaborative manner, where students and AI work together. It can be used as a teammate, tutor, or a resource to support learning. It is not simply a cheating tool.	9
<i>Learning Experience</i>	Participants found the assignment to be a valuable learning experience. They learned about the subject and ChatGPT's capabilities, how to use it effectively, and its limitations.	4	<i>Appropriate Reliance</i>	ChatGPT should not be relied upon entirely. It should be used as a reference, tool, or method to complement human thought, but not generate whole papers. There are limitations and assignments like this help identify the areas where ChatGPT falls short	5
<i>Need for Manual Writing</i>	Some participants noticed that ChatGPT's first drafts and responses were not up to the desired standard, requiring extensive additional work, analysis, and editing.	3	<i>Error Checking and Auto Grading</i>	A few students identified the potential for errors in code and autograding, although were explicit in not trusting ChatGPT to grade without human oversight	3
<i>Skepticism and Limited Trust</i>	Some participants remained skeptical about ChatGPT's capabilities and were reluctant to fully rely on its responses.	3	<i>Redefining Education</i>	The technology could redefine education, and its full impact is uncertain, although this should be implemented with caution.	3

who described it as the “coolest assignment” they had encountered, expressing their belief in the transformative potential of AI tools, such as ChatGPT for the future:

“Coolest assignment I’ve done to date. I think tools like ChatGPT will change our future and assignments like these are paramount to understanding the direction we want to take them.”

On the other hand, it was apparent that the cadets’ views on the assignment’s learning value were more varied and some expressed the shift in the skills being assessed:

“I thought it was more of an assessment of our editing skills than our opinions on the topic.”

While the cadets acknowledged the exposure to cutting-edge technology, concerns emerged regarding the assignment’s applicability to assessments and its potential for grading. One cadet noted that, while they found the experience valuable, they were not comfortable with ChatGPT being involved in the grading process:

“I thought it was a great exposure to the technology that we are going to be seeing so much more of in the future! I recommend keeping this assignment, but I don’t recommend having ChatGPT grade it.”

This apprehension seemed to indicate a lack of trust in the AIs ability to provide reliable and accurate assessments. Students’ open-ended responses also uncovered cadets’ overall trustworthiness and comfort level when working with ChatGPT. Several participants expressed feelings of discomfort, likely stemming from the assignment’s departure from their norms. Wrestling with ChatGPT’s responses and seeking ways to integrate them effectively highlighted the cadets’ commitment to taking responsibility for their work:

“I think it was an interesting assignment, although I felt a little uncomfortable doing it just because it was outside of my norm. I found myself wrestling with ChatGPT on some parts but I feel like there were times it provided some decent feedback that I could improve on. In my mind it is still nothing more than a tool and I find it difficult to rely wholly upon it.”

IV. DISCUSSION

In this study, we explored student perceptions of ChatGPT in the context of essay writing within an engineering course focusing on three RQs tied to the following:

- 1) ChatGPT use in an assignment;
- 2) its capabilities to support learning;

- 3) student trust and comfort in relying on ChatGPT for future assignments and grading.

Our results showed that ChatGPT did not make the writing assignment easy but changed it in ways that yielded perceived learning benefits for students. The thematic analysis revealed a shift in student perception, evolving from viewing ChatGPT as a potential “cheating tool” to recognizing it as a collaborative resource requiring human oversight, technical aptitude, subject-area proficiency, and calibrated trust. After use, students rated ChatGPT as a valuable tool for learning and more ethical and benevolent relative to their perceptions before use. Students’ low comfort in taking responsibility for the assignment could be attributed to ethical concerns, given the high percentage of students who believe using ChatGPT is akin to cheating [12]. Additionally, the lack of full confidence in the accuracy and reliability of ChatGPT’s output likely contributed to students’ discomfort after the assignment. These findings highlight the complex dynamics associated with integrating AI tools, such as ChatGPT, into higher education.

Students did not want to be evaluated on this assignment by ChatGPT alone, instead preferring to be graded by ChatGPT and the instructor together or by the instructor alone. Overall, our results reveal that technologies, such as ChatGPT, do not eliminate the need for student and instructor engagement, but rather complement it, requiring judicious trust and a blend of human skill and AI capabilities. In educational contexts, this integration of AI tools with student participation seemed to foster an effective learning experience according to students, yet also revealed areas where ChatGPT and its integration could be improved.

A. Implications for Student Learning

STEM and non-STEM educators should be encouraged to integrate AI technologies in deliberate ways to promote student learning, with some caution. An assignment requiring the use of ChatGPT to produce better papers was widely accepted by students as valuable. Some students were enthusiastic, describing it as the “coolest assignment” they had encountered, emphasizing the transformative potential of AI tools, such as ChatGPT for the future. The assignment was also viewed by students as difficult and more difficult after they used ChatGPT to complete it. While some of this was due to usability and related concerns with ChatGPT, the assignment also required “more” from our students: both in terms of the number of drafts required to turn in (i.e., three versions of the paper) and the overall quality expected. ChatGPT helped them reach higher levels as students’ self-reported assessment of their paper quality was high. Despite this result, grades for this assignment were comparable to grades for similar essay assignments that did not mandate the use of ChatGPT. It is possible that the beneficial or detrimental effects of ChatGPT are more subtle, such as the finding that the tool is particularly beneficial for weaker performers [87].

While students recognized the learning value and improved essay quality facilitated by ChatGPT’s feedback, they also grappled with calibration issues between their initial expectations and the actual outcomes. This could be a double-edged sword

for educators. On the one hand, engaged learning is a cornerstone of successful knowledge and skill acquisition [62], [63], [64], [65], and reviewing, critiquing, and editing papers are effective for engaged learning [66]. The more critical students are in their reviews of others’ papers, the better they do on their own work, and this leads to better learning of writing skills and more knowledge of the subject material itself [66]. Reasons for this include that their reading is not passive, but active and critical to identify the aspects of writing to keep, but also elements they want to avoid (e.g., poor writing structure and mistakes). Like assignments requiring students to critique their peers, the essay assignment used in this study required participants to critique ChatGPT output. This active reviewing, editing, and writing are integrated to reinforce the other and this is invaluable to student learning [67]. Based on our results, students recognized the learning benefits of integrating AI tools, such as ChatGPT, in educational settings.

On the other hand, some of the student engagement with ChatGPT was frustrating and did not seem beneficial for learning. Recall the theme ChatGPT produced in its thematic analysis of students’ open-ended responses:

“ChatGPT performed poorly on the assignment and was often very repetitive. It was difficult to use and did not do a good job.”

Participants mentioned challenges in creating prompts, managing word counts, and dealing with repetitive language. These findings indicate the need for clear practical guidelines and instructions for students to optimize their interactions with AI tools and streamline the assignment process, which instructors should provide. Guidance and training on creating suitable prompts, for example, can help students optimize their use of AI tools and reduce potential frustrations.

B. Design Recommendations for LLMs in Education

The findings also point to the further development and customization of LLMs for educational contexts. AI developers should continue refining and improving language models, such as ChatGPT, to enhance their effectiveness and reliability. Common issues, such as repetitive language generation and inadequate essay production, should be addressed to ensure a more seamless and valuable user experience. Furthermore, metadata could be provided about the tool’s confidence to assess whether it is producing the accurate output for a specific prompt, by providing a hallucinating score, for example. Such confidence indicators have been known to increase trust calibration [68], [69] and help to increase the transparency of the tool [70]. ChatGPT was not designed specifically for student learning. As LLM capabilities are enhanced and more customized AI tools are developed for education, it is likely to be better for more tailored student engagement.

C. Improving the Writing Process With ChatGPT

Instructors should further experiment with integrating ChatGPT intentionally [71] versus open-ended use by considering the iterative and dynamic nature of the writing process, as well as the

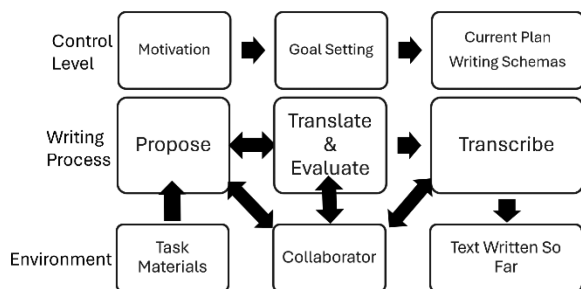


Fig. 4. Adapted from Hayes' (2012) model of writing [75]. ChatGPT as a "Collaborator" helps "propose" by generating ideas, "translate" by converting concepts to text, "evaluate" by providing feedback, and "transcribe" by drafting content. This new process likely impacts the traditional human writing schemas.

specific functions that need to be performed to accomplish a writing assignment (see Fig. 4). We deliberately designed the task to tradeoff text generation by ChatGPT, editing by the human, and then combining the product. By design, combined human and AI teaming was necessary for the successful completion of the task [72], as is mostly the case with the human use of automated tools [73] and human–autonomy teaming [74]. This approach is often advocated when automation is not perfect, and interdependence is required to create good team performance [68]. Guiding students to collaborate with the AI, instead of preventing its use, or allowing supervised use without specific constraints could be beneficial to discover the best way to integrate ChatGPT into course assignments.

It is not clear yet where ChatGPT can be most effective in the writing process when working collaboratively with people, as opposed to strictly working by itself, but our data suggest some possible directions. Students indicated it could be useful in the early stages of idea generation, providing topic information, producing text from rough ideas, and reviewing the text. Other potential uses could be in specific roles at the process level (proposing, evaluating, and transcribing) or providing high-level writing schemas and monitoring them at the control level (see Fig. 4 [75]). Students can further be encouraged to use AI-generated content as a starting point and then iteratively refine and enhance their essays. Others have suggested using ChatGPT to prepare outlines, revise content, proofread the paper, or reflect on the writing [71]. This approach can foster a deeper understanding of the writing process and enable students to develop their skills through active engagement with AI technology. Traditional schools and platforms, such as Khan Academy and Udacity, are increasingly exploring the integration of AI tools, such as ChatGPT, to enhance personalized learning experiences and to supplement their existing course materials.

As discussed above, even simple integration of AI tools to encourage exchanges with ChatGPT likely yielded valuable repetition to practice the skill of writing in an iterative way. One study examined the performance of over 4000 students using project essay grading automated grading for writing and revising essays with feedback [76]. Students who revised their papers achieved small score increases with each draft, although the rate of growth diminished over time, reaching a plateau around the 11th or 12th revision. LLMs offer the potential to

provide more precise and effective feedback, perhaps requiring fewer revisions because of enhanced feedback; however, more research is needed to determine the differences between other ways of learning to write versus a more cyclic ChatGPT-writing approach. The taxonomy for levels and degrees of automation could be helpful as an initial guide [77], [78], [79] to distinguish between the use of AI for initial ideas and inspiration (low writing automation) to using AI for early drafts, feedback, and edits (medium writing automation) and purely AI-generated work with little human input.

D. Implications for Grading

Most students preferred the instructor to use ChatGPT to support grading and fewer students preferred the teacher alone or the LLM alone. Before ChatGPT, AWE technologies were developed to help teachers save time in assessing writing, encourage more writing practice, and complement writing instruction in the classroom. Similar to AWEs, ChatGPT can assess essays in seconds and enable teachers to assign more writing tasks without an overwhelming increase in workload. Still, students expressed the importance of human oversight: one of the strongest effects of this study was that no student wanted to be graded by ChatGPT alone. This result appears to be consistent with a preference to receive writing feedback from teachers or peers rather than computers [30], [31], [32]. The strong reluctance to rely on ChatGPT alone for grading suggests the need for further investigation into the factors influencing students' trust and confidence in AI-generated outputs, particularly when they bear personal responsibility for the final work. This also raises questions about the role of AI in the evaluation of student work. It seemed that when the stakes were high (e.g., assigning official grades), human instructor involvement was perceived as critical. Despite this preference, and the high trust in the instructor alone, most students preferred to be graded by the instructor and ChatGPT together. It is possible that students thought that some involvement of ChatGPT, as an assistive tool perhaps, in grading would be beneficial rather than the instructor grading by himself. This result speaks to a potential required shift in work for both teachers and students in academic settings.

E. Implications for Trust in AI

Our study is among the first to rigorously explore trust in AI, using ChatGPT, within the education domain extending the existing trust models, originally developed for human–automation interaction, to a novel setting [43], [50], [80]. Our regression model showed strong support for existing trust models and previous research validating our theoretical and measurement framework and investigations of trust with ChatGPT specifically [43], [48], [81], [86]. Moreover, our study extends our understanding of trust in AI by highlighting the pivotal role of moral trust, particularly in the realms of ethical and benevolent trustworthiness, in shaping the overall trust and reliance on AI systems, such as ChatGPT. The increase in these dimensions after using ChatGPT (with medium to large effect sizes) was novel. It is possible that students perceived ChatGPT to be more ethically trustworthy because they received content violation

messages while using the program, indicating that ChatGPT tries to adhere to OpenAI's content policy. Such behavior may have increased perceptions of the moral competence of ChatGPT, which can increase trustworthiness [42]. Ethicality may have further increased because students perceived ChatGPT as providing information without personal beliefs or opinions. Benevolence trustworthiness may have increased because students observed ChatGPT's behavior to be very helpful and responsive to feedback from the user. While ChatGPT was considered trustworthy overall, intent to rely on ChatGPT was lower, which is consistent with findings from previous work [42], [85]. This result suggests that students understood ChatGPT's capabilities but also realized that they could not fully depend on it alone to complete the assignment; a conclusion consistent with the decades of human factors automation research that demonstrates automation does not replace the human but changes the way we work with technology [15], [77].

F. Limitations

There were a few limitations to this study. Most notably, the sample size was relatively small taking the advantage of timing, where ChatGPT had not been widely used by students yet and certainly not incorporated into curricula. Low sample sizes are not uncommon with early studies on technology integration, including for ChatGPT [8], smartphone use [82], and robots [83], presenting a tradeoff between the impact of novel technology use and the generalizability of results. However, we do believe that our sample was representative of senior undergraduate engineering students as well as novice scientific writers. The use of Bayes factors uniquely accommodates small samples, and the moderate-to-large effect sizes add confidence to our results. However, our sample size may have precluded us from identifying smaller effects. Future studies should assess an increased number of students. In addition, long-term investigations that track students' experiences and perceptions of AI-powered tools could provide valuable insights into the evolving dynamics between students and AI technology and identify areas for continuous improvement.

Second, the assignment was highly structured and specified how students had to work with ChatGPT in three iterative steps to show the contributions of the AI versus the student. However, this may have constrained the use of ChatGPT in other more creative ways or in ways that suited the student better. Another potential approach would have been to have the student use the tool in anyway they like, but acknowledge in the final product where ChatGPT had assisted, as we have in this paper for summarization of qualitative results and in another for figure generation [42]. Students could either summarize ChatGPT assistance in the acknowledgment section or provide brief annotations on paper in the places where ChatGPT specifically assisted. There are many ways to incorporate ChatGPT in an assignment and our approach represented one way. As such, our results could reflect our design of the assignment and not ChatGPT or other LLM capabilities more broadly. We hope this report encourages the novel uses of LLM technologies in course didactics and future studies of their effectiveness.

Finally, the participants in this study were also senior-level undergraduate human systems engineering students. Their education in human factors processes, including knowledge elicitation through survey-based user feedback methods (the approach used in this study), could have influenced their responses.

G. Contributions

These limitations notwithstanding, our study provides empirically grounded insights that are important for understanding how students perceive and interact with AI in educational settings. Our study contributes one of the first mixed-method studies exploring ChatGPT's application in college classes informing research and design vis-à-vis intelligent technology use in naturalistic contexts. To this end, we used a unique approach to formally integrate ChatGPT as a part of a writing curriculum in the semester following the worldwide release of ChatGPT. Our study provided a common AI interaction for engineering students—before LLMs became ubiquitous—allowing a focused analysis of how such exposure specifically influences their perceptions and attitudes toward AI in essay writing tasks. This included writing a full technical paper, not just a few paragraphs, and had real consequences for students who received an actual grade for their assignment. ChatGPT was evaluated as a writing assistant as well as a grading assistant with implications for theory, design, and practice.

Beyond pre- and post-survey perceptions of ChatGPT use for learning, grading, and their comfort in taking responsibility for the essay produced with its help, this report provides the first comprehensive trust assessment for ChatGPT, which has strong validity due to the real vulnerability and consequences associated with this assignment; a requirement for accurate trust assessment [43], [48], [80]. Theoretically, this study extends established trust models to a new domain, education, and to understand how students build trust in AI technologies, such as ChatGPT, in a learning environment. Practically, this study demonstrates the feasibility of integrating AI-based technologies into the classroom and provides usability insights and recommendations based on user (i.e., student) feedback. By integrating approaches from both human factors engineering and educational research, our study not only provides practical insights for the design of AI-enhanced educational tools but also a multidisciplinary examination of the complex dynamics of trust in AI systems.

V. CONCLUSION

Our research indicates that, while AI tools, such as ChatGPT, have promising applications in higher education, they also pose challenges. These tools should be viewed as helpful assistants to enhance writing, learning, and grading and not as replacements for student effort or teaching oversight in grading. Effective and ethical use of AI in education requires acknowledging its limitations, fostering AI literacy, and developing proper assessment methods. Institutions must also train students and educators in using AI responsibly and creatively. Future studies should focus on improving understanding of AI, guiding its use, and addressing issues with AI-generated content.

APPENDIX

Table III describes all the measures that were included in the surveys created for this study.

TABLE III
CUSTOM ITEM SURVEY MEASURES

Assignment
1. How would you rate the difficulty of this assignment on a scale of 1 (Very easy) to 7 (Very difficult).
2. How would you rate the difficulty of this assignment relative to writing the paper by yourself (with no help from ChatGPT) on a scale of 1 (Very easy) to 7 (Very difficult).
3. Estimate what grade percentage (1%–100%) you believe you will obtain for this assignment.
Learning
1. Rate how valuable of a learning experience you think this will be/was on a scale of 1 (Not very valuable) to 7 (Very valuable).
2. Rate how valuable of a learning experience you think this will be/was relative to an essay you have already written on a scale of 1 (Not very valuable) to 7 (Very valuable).
3. Relative to an essay you have already written, rate what you think the quality of this paper will be/was on a scale of 1 (Lower) to 7 (Higher).
4. What is / was your comfort level being responsible for all content created by ChatGPT on a scale of 1 (Very uncomfortable) to 7 (Very comfortable).
5. Generally, how did you feel about this assignment (open response)?
6. How well has this technology (e.g. ChatGPT) been integrated into college education (open response)?
7. How should this technology (e.g. ChatGPT) be integrated into college education (open response)?
Grading
1. I trust that the instructor will grade fairly on his own from 0 (Strongly Disagree) to 7 (strongly agree).
2. I trust that ChatGPT will grade fairly on it from 0 (Strongly Disagree) to 7 (strongly agree).
3. I trust that the instructor and ChatGPT will grade fairly together from 0 (Strongly Disagree) to 7 (strongly agree).
4. Of the three options (the instructor on his own, ChatGPT on its own, or the instructor and ChatGPT together), my preference is for _____ to grade my paper from 0 (Strongly Disagree) to 7 (Strongly Agree).

ACKNOWLEDGMENT

The views expressed in this article are those of the authors and do not reflect those of the U.S. Air Force, DARPA, Department of Defense, or U.S. Government. Please contact the authors for access to the assignment and/or rubric used in this study.

REFERENCES

- [1] "OpenAI releases ChatGPT: An advanced AI language model for text-based interactions." 2023. [Online]. Available: <https://openai.com/chatgpt>
- [2] N. Manohar and S. S. Prasad, "Use of ChatGPT in academic publishing: A rare case of seronegative systemic lupus erythematosus in a patient with HIV infection," *Cureus*, vol. 15, no. 2, Feb. 2023, Art. no. e34616, doi: [10.7759/cureus.34616](https://doi.org/10.7759/cureus.34616).
- [3] D. Baidoo-Anu and L. Owusu Anash, "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning," *J. AI*, Jan. 2023. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4337484
- [4] W. Heaven, "ChatGPT is going to change education, not destroy it," *MIT Technol. Rev., Artif. Intell.*, Apr. 2023. [Online]. Available: <https://www.technologyreview.com/2023/04/06/1071059/chatgpt-change-not-destroy-education-openai/>
- [5] M. Sullivan, A. Kelly, and P. McLaughlan, "ChatGPT in higher education: Considerations for academic integrity and student learning," *J. Appl. Learn. Teach.*, vol. 6, no. 1, Mar. 2023, doi: [10.37074/jalt.2023.6.1.17](https://doi.org/10.37074/jalt.2023.6.1.17). [Online]. Available: https://www.researchgate.net/publication/369378950_ChatGPT_in_higher_education_Considerations_for_academic_integrity_and_student_learning
- [6] S. Marche, "The college essay is dead," *The Atlantic*, Dec. 2022. [Online]. Available: <https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/>
- [7] L. Lonas, "Professor attempts to fail entire after falsely accusing them of using ChatGPT to cheat," *The Hill*, May 2023. [Online]. Available: <https://thehill.com/homenews/education/4010647-professor-attempts-to-fail-students-after-falsely-accusing-them-of-using-chatgpt-to-cheat/>
- [8] M. Firat, "How chat GPT can transform autodidactic experiences and open education?," Jan. 2023, doi: [10.31219/osf.io/9ge8m](https://doi.org/10.31219/osf.io/9ge8m). [Online]. Available: https://www.researchgate.net/publication/367613715_How_Chat_GPT_Can_Transform_Autodidactic_Experiences_and_Open_Education
- [9] M. Firat, "What ChatGPT means for universities: Perceptions of scholars and students," *J. Appl. Learn. Teach.*, vol. 6, no. 1, Apr. 2023, doi: [10.37074/jalt.2023.6.1.22](https://doi.org/10.37074/jalt.2023.6.1.22). [Online]. Available: https://www.researchgate.net/publication/370107010_What_ChatGPT_means_for_universities_Perceptions_of_scholars_and_students
- [10] M. M. Rahman and Y. Watanobe, "ChatGPT for education and research: Opportunities, threats, and strategies," *Appl. Sci.*, vol. 13, May 2023, Art. no. 5783, doi: [10.3390/app13095783](https://doi.org/10.3390/app13095783). [Online]. Available: <https://www.mdpi.com/2076-3417/13/9/5783>
- [11] S. Stacey, "Cheating on your college essay with ChatGPT won't get you good grades, say professors—But AI could make education fairer," *MSN*, Dec. 2022. [Online]. Available: <https://www.msn.com/en-us/news/technology/cheating-on-your-college-essay-with-chatgpt-wont-get-you-good-grades-say-professors-but-ai-could-make-education-fairer/ar-AA15pCJ8?li=BBnbcA1>
- [12] M. Nietzel, "More than half of college students believe using ChatGPT to complete assignments is cheating," *Forbes*, Mar. 2023. [Online]. Available: <https://www.forbes.com/sites/michaelnietzel/2023/03/20/more-than-half-of-college-students-believe-using-chatgpt-to-complete-assignments-is-cheating/?sh=5d4d763c18f9>
- [13] R. Ventayen, "ChatGPT by OpenAI: Student's viewpoint on cheating using artificial intelligence-based application," *SSRN*, Feb. 2023, doi: [10.2139/ssrn.4361548](https://doi.org/10.2139/ssrn.4361548). [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4361548#:~:text=Abstract,the%20integrity%20of%20student%20submissions
- [14] A. Shoufan, "Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey," *IEEE Access*, vol. 11, pp. 38805–38818, 2023, doi: [10.1109/ACCESS.2023.3268224](https://doi.org/10.1109/ACCESS.2023.3268224).
- [15] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *J. Hum. Factors Ergonom. Soc.*, vol. 39, no. 2, Jun. 1997, doi: [10.1518/001872097778543886](https://doi.org/10.1518/001872097778543886). [Online]. Available: <http://web.mit.edu/16.459/www/parasuraman.pdf>
- [16] K. Muenks et al., "Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset beliefs predict their psychological vulnerability, engagement, and performance in class," *J. Exp. Psychol. Gen.*, vol. 149, no. 11, pp. 2119–2144, May 2020, doi: [10.1037/xge0000763](https://doi.org/10.1037/xge0000763).
- [17] J. Van Brummelen, V. Tabunshchik, and T. Heng, "Alexa, can I program you?: Student perceptions of conversational artificial intelligence before and after programming Alexa," in *Proc. 20th Annu. ACM Interaction Des. Child. Conf.*, 2021, pp. 305–313, doi: [10.1145/3459990.3460730](https://doi.org/10.1145/3459990.3460730).
- [18] C. Zhang et al., "One small step for generative AI, one giant leap for AGI: A complete survey on chatGPT in AIGC era," *ACM*, Mar. 2023, doi: [10.13140/RG.2.2.24789.70883](https://doi.org/10.13140/RG.2.2.24789.70883). [Online]. Available: https://www.researchgate.net/publication/369618942_One_Small_Step_for_Generative_AI_One_Giant_Leap_for_AGI_A_Complete_Survey_on_ChatGPT_in_AIGC_Era
- [19] O. Topsakal and E. Topsakal, "Framework for a foreign language teaching software for children utilizing AR, Voicebots and ChatGPT (large language models)," *J. Cogn. Syst.*, vol. 7, no. 2, pp. 33–38, 2022, doi: [10.52876/jcs.1227392](https://doi.org/10.52876/jcs.1227392).
- [20] W. Yeadon, O. O. Inyang, A. Mizouri, A. Peach, and C. P. Testrow, "The death of the short-form physics essay in the coming AI revolution," *Phys. Educ.*, vol. 58, no. 3, pp. 1–13, Dec. 2022, doi: [10.48550/arXiv.2212.11661](https://doi.org/10.48550/arXiv.2212.11661).
- [21] S. Frieder et al., "Mathematical capabilities of chatGPT," Jan. 2023, *arXiv:2301.13867*.
- [22] J. Qadir, "Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education," in *Proc. IEEE Glob. Eng. Educ. Conf.*, 2023, pp. 1–9.
- [23] B. Warner, "ChatGPT is and is not like calculators," *Inside Higher Educ.*, Accessed on: Jul. 26, 2023. [Online]. Available: <https://www.insidehighered.com/blogs/just-visiting/chatgpt-both-and-not-calculator>

- [24] Y. K. Dwivedi et al., "Opinion paper: 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *Int. J. Inf. Manage.*, vol. 71, Aug. 2023, doi: [10.1016/j.ijinfomgt.2023.102642](https://doi.org/10.1016/j.ijinfomgt.2023.102642). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0268401223000233>
- [25] J. Rudolph, S. Tan, and S. Tan, "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?," *J. Appl. Learn. Teach.*, vol. 6, no. 1, pp. 342–363, 2023, doi: [10.37074/jalt.2023.6.1.9](https://doi.org/10.37074/jalt.2023.6.1.9). [Online]. Available: <https://journals.sfu.ca/jalt/index.php/jalt/article/view/689>
- [26] E. Glikson and A. W. Woolley, "Human trust in artificial intelligence: Review of empirical research," *Acad. Manage. Ann.*, vol. 14, no. 2, pp. 627–660, Aug. 2020, doi: [10.5465/annals.2018.0057](https://doi.org/10.5465/annals.2018.0057).
- [27] A. D. Kaplan, T. T. Kessler, J. C. Brill, and P. A. Hancock, "Trust in artificial intelligence: Meta-analytic findings," *Hum. Factors*, vol. 65, no. 2, pp. 337–359, May 2021, doi: [10.1177/00187208211013988](https://doi.org/10.1177/00187208211013988).
- [28] D. Grimes and M. Warshauer, "Utility in a fallible tool: A multi-site case study of automated writing evaluation," *J. Technol., Learn. Assessment*, vol. 8, no. 6, Mar. 2010. [Online]. Available: <https://ejournals.bc.edu/index.php/jtla/article/view/1625>
- [29] M. Warschauer and D. Grimes, "Automated writing assessment in the classroom," *Pedagogies, Int. J.*, vol. 3, no. 1, Jan. 2008, doi: [10.1080/15544800701771580](https://doi.org/10.1080/15544800701771580). [Online]. Available: <https://education.uci.edu/uploads/7/12/7/6/72769947/awe-pedagogies.pdf>
- [30] P. Haber-Curran, D. Tillapaugh, and M. Orono, "Leadership learning through student-centered and inquiry-focused approaches to teaching adaptive leadership," *J. Leadership Educ.*, vol. 12, no. 1, pp. 92–116, 2013, doi: [10.12806/V12/I1/R6](https://doi.org/10.12806/V12/I1/R6).
- [31] Y.-H. Lai, "Which do students prefer to evaluate their essays: Peers or computer program," *Brit. J. Educ. Technol.*, vol. 41, Apr. 2010, doi: [10.1111/j.1467-8535.2009.00959.x](https://doi.org/10.1111/j.1467-8535.2009.00959.x). [Online]. Available: <https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8535.2009.00959.x>
- [32] A. A. Lipnevich and J. K. Smith, "Effects of differential feedback on students' examination performance," *J. Exp. Psychol. Appl.*, vol. 15, no. 4, pp. 319–333, 2009, doi: [10.1037/a0017841](https://doi.org/10.1037/a0017841).
- [33] B. M. Olds, B. M. Moskal, and R. L. Miller, "Assessment in engineering education: Evolution, approaches and future collaborations," *J. Eng. Educ.*, vol. 94, pp. 13–25, Jan. 2005, doi: [10.1002/j.2168-9830.2005.tb00826.x](https://doi.org/10.1002/j.2168-9830.2005.tb00826.x).
- [34] C. Teddlie and A. Tashakkori, "Major issues and controversies in the use of mixed methods in the social and behavioural sciences," in *Handbook of Mixed Methods in Social and Behavioural Research*. Newbury Park, CA, USA: Sage, 2003, pp. 3–50.
- [35] K. Gogol et al., "My questionnaire is too long! The assessments of motivational-affective constructs with three-item and single-item measures," *Contemporary Educ. Psychol.*, vol. 39, no. 3, pp. 188–205, Jul. 2014, doi: [10.1016/j.cedpsych.2014.04.002](https://doi.org/10.1016/j.cedpsych.2014.04.002).
- [36] P. Lukowicz, A. Choynowska, A. M. Swiatkowska, and P. Bereznowski, "Validity of single-item self-report measure of learning engagement," in *Badania i Rozwój Młodych Naukowców w Polsce – Nauki Humanistyczne i Społeczne*. Poznan, Poland: Młodzi Naukowcy, 2006, pp. 41–49.
- [37] A. Bartoszko, S. Czerwinski, and A. Bereznowska, "Single-item self-report measure of learning engagement: What does it measure?," presented at Interdisciplinary Scientific International Conference for PhD Students and Assistants, Jun. 2019. [Online]. Available: https://www.researchgate.net/publication/334974047_Single-item_self-report_measure_of_learning_engagement_What_does_it_measure
- [38] B. Barker and D. Brooks, "An evaluation of short-term distributed online learning events," *Int. J. E-Learn.*, vol. 4, no. 2, pp. 209–228, Jun. 2005.
- [39] R. A. Wisher and C. K. Curnow, "Techniques for evaluating distance learning events," U.S. Army Res. Inst., Alexandria, VA, USA, Tech. Rep. IR019028, 1998.
- [40] D. Ullman and B. F. Malle, "Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust," in *Proc. 14th ACM/IEEE Int. Conf. Human-Robot Interaction*, 2019, pp. 618–619.
- [41] B. F. Malle and D. Ullman, "A multidimensional conception and measure of human-robot trust," in *Trust in Human-Robot Interaction*. New York, NY, USA: Academic, 2021, pp. 3–25.
- [42] A. Momen, E. De Visser, K. Wolsten, K. Cooley, J. Walliser, and C. C. Tossell, "Trusting the moral judgments of a robot: Perceived moral competence and humanlikeness of a GPT-3 enabled AI," in *Proc. 56th Hawaii Int. Conf. Syst. Sci.*, 2023, pp. 501–510.
- [43] S. C. Kohn, E. J. de Visser, E. Wiese, Y.-C. Lee, and T. H. Shaw, "Measurement of trust in automation: A narrative review and reference guide," *Front. Psychol.*, vol. 12, Oct. 2021, Art. no. 604977, doi: [10.3389/fpsyg.2021.604977](https://doi.org/10.3389/fpsyg.2021.604977).
- [44] S. S. Monfort, J. J. Graybeal, A. E. Harwood, P. E. McKnight, and T. H. Shaw, "A single-item assessment for remaining mental resources: Development and validation of the gas tank questionnaire (GTQ)," *Theor. Issues Ergonom. Sci.*, vol. 19, no. 5, pp. 530–552, 2018, doi: [10.1080/1463922X.2017.1397228](https://doi.org/10.1080/1463922X.2017.1397228).
- [45] J. B. Lyons and S. Y. Guznov, "Individual differences in human-machine trust: A multi-study look at the perfect automation schema," *Theor. Issues Ergonom. Sci.*, vol. 20, no. 4, pp. 440–458, Nov. 2018, doi: [10.1080/1463922X.2018.1491071](https://doi.org/10.1080/1463922X.2018.1491071).
- [46] M. T. Dzindolet, S. A. Peterson, R. A. Pomranky, L. G. Pierce, and H. P. Beck, "The role of trust in automation reliance," *Int. J. Human-Comput. Stud.*, vol. 58, no. 6, pp. 697–718, Jun. 2003, doi: [10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7).
- [47] N. L. Tenhundfeld, E. J. de Visser, A. J. Ries, V. S. Finomore, and C. C. Tossell, "Trust and distrust of automated parking in a Tesla Model X," *Hum. Factors*, vol. 62, no. 2, pp. 194–210, Aug. 2019, doi: [10.1177/0018720819865412](https://doi.org/10.1177/0018720819865412).
- [48] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An integrative model of organizational trust," *Acad. Manage. Rev.*, vol. 20, no. 3, pp. 709–734, Jul. 1995, doi: [10.2307/258792](https://doi.org/10.2307/258792).
- [49] S. M. Merritt and D. R. Ilgen, "Not all trust is created equal: Dispositional and history-based trust in human-automation interactions," *Hum. Factors*, vol. 50, no. 2, pp. 194–208, Apr. 2008, doi: [10.1518/001872008X288574](https://doi.org/10.1518/001872008X288574).
- [50] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Hum. Factors*, vol. 57, no. 3, pp. 407–434, Sep. 2014, doi: [10.1177/0018720814547570](https://doi.org/10.1177/0018720814547570).
- [51] I. L. Singh, R. Molloy, and R. Parasuraman, "Automation-induced 'complacency': Development of the complacency-potential rating scale," *Int. J. Aviation Psychol.*, vol. 3, no. 2, pp. 111–122, 1993, doi: [10.1207/s15327108ijap0302_2](https://doi.org/10.1207/s15327108ijap0302_2).
- [52] S. M. Merritt et al., "Automation-induced complacency potential: Development and validation of a new scale," *Front. Psychol.*, vol. 10, Feb. 2019, Art. no. 225, doi: [10.3389/fpsyg.2019.00225](https://doi.org/10.3389/fpsyg.2019.00225).
- [53] S. M. Merritt, H. Heimbaugh, J. LaChapell, and D. Lee, "I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system," *Hum. Factors*, vol. 55, no. 3, pp. 520–534, Jun. 2013, doi: [10.1177/0018720812465081](https://doi.org/10.1177/0018720812465081).
- [54] D. Schmid and N. A. Stanton, "Exploring Bayesian analyses of a small-sample-size factorial design in human systems integration: The effects of pilot incapacitation," *Human-Intell. Syst. Integr.*, vol. 1, pp. 71–88, 2019, doi: [10.1007/s42454-020-00012-0](https://doi.org/10.1007/s42454-020-00012-0).
- [55] E.-J. Wagenmakers et al., "Bayesian inference for psychology—Part I: Theoretical advantages and practical ramifications," *Psychon. Bull. Rev.*, vol. 25, pp. 35–57, Feb. 2018, doi: [10.3758/s13423-017-1343-3](https://doi.org/10.3758/s13423-017-1343-3).
- [56] S. Andraszewicz, B. Scheibehenne, J. Rieskamp, R. Grasman, J. Verhagen, and E.-J. Wagenmakers, "An introduction to Bayesian hypothesis testing for management research," *J. Manage.*, vol. 41, no. 2, pp. 521–543, 2015, doi: [10.1177/0149206314560412](https://doi.org/10.1177/0149206314560412).
- [57] R. Kelter, "Bayesian alternatives to null hypothesis significance testing in biomedical research: A non-technical introduction to Bayesian inference with JASP," *BMC Med. Res. Methodol.*, vol. 20, no. 1, Jun. 2020, Art. no. 142, doi: [10.1186/s12874-020-00980-6](https://doi.org/10.1186/s12874-020-00980-6).
- [58] J. Cohen, "A power primer," *Psychol. Bull.*, vol. 112, no. 1, pp. 155–159, Jul. 1992, doi: [10.1037//0033-2909.112.1.155](https://doi.org/10.1037//0033-2909.112.1.155).
- [59] S. Bubeck et al., "Sparks of artificial general intelligence: Early experiments with GPT-4," *Microsoft Res.*, Apr. 2023, doi: [10.48550/arXiv.2303.12712](https://doi.org/10.48550/arXiv.2303.12712).
- [60] J. Yang et al., "Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond," Apr. 2023, *arXiv:2304.13712*.
- [61] P. Mackieson, A. Shlonsky, and M. Connolly, "Increasing rigor and reducing bias in qualitative research: A document analysis of parliamentary debates using applied thematic analysis," *Qualitative Social Work*, vol. 18, no. 6, pp. 965–980, 2019, doi: [10.1177/1473325018786996](https://doi.org/10.1177/1473325018786996).
- [62] T. V. McGovern, "Process/pedagogy," in *The Teaching of Psychology: Essays in Honors of Wilbert J. McKeachie and Charles L. Brewer*. Mahwah, NJ, USA: Lawrence Erlbaum Associates Inc., 2002, pp. 81–90.
- [63] W. J. McKeachie, "Teaching, learning, and thinking about teaching and learning," in *Higher Education: Handbook of Theory and Research*, vol. 14. Dordrecht, The Netherlands: Springer, 1999, pp. 1–38.

- [64] W. J. McKeachie, "Ebbs, flows, and progress in the teaching of psychology," in *The Teaching of Psychology: Essays in Honors of Wilkert J. McKeachie and Charles L. Brewer*. Mahwah, NJ, USA: Lawrence Erlbaum Associates Inc, 2002, pp. 487–498.
- [65] M. Miserandino, "Those who can do: Implementing active learning," in *Lessons Learned: Practice Advice for the Teaching of Psychology*. Washington, DC, USA, American Psychological Society, 1999, pp. 109–114.
- [66] M. M. Yalch, E. M. Vitale, and J. K. Ford, "Benefits of peer review on students' writing," *Psychol. Learn. Teach.*, vol. 18, no. 3, pp. 317–325, Apr. 2019, doi: [10.1177/1475725719835070](https://doi.org/10.1177/1475725719835070).
- [67] P. Elbow, *Everyone Can Write: Essays Toward a Hopeful Theory of Writing and Teaching Writing*. London, U.K.: Oxford Univ. Press, 2000.
- [68] E. de Visser and R. Parasuraman, "Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload," *J. Cogn. Eng. Decis. Mak.*, vol. 5, no. 2, pp. 209–231, Jun. 2011, doi: [10.1177/1555343411410160](https://doi.org/10.1177/1555343411410160).
- [69] J. M. McGuirl and N. B. Sarter, "Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information," *Hum. Factors*, vol. 48, no. 4, pp. 656–665, 2006, doi: [10.1518/001872006779166334](https://doi.org/10.1518/001872006779166334).
- [70] J. Y. C. Chen, S. G. Lakhmani, K. Stowers, A. R. Selkowitz, J. L. Wright, and M. Barnes, "Situation awareness-based agent transparency and human-autonomy teaming effectiveness," *Theor. Issues Ergonom. Sci.*, vol. 19, no. 3, pp. 259–282, Feb. 2018, doi: [10.1080/1463922X.2017.1315750](https://doi.org/10.1080/1463922X.2017.1315750).
- [71] Y. Su, Y. Lin, and C. Lai, "Collaborating with ChatGPT in argumentative writing classrooms," *Accessing Writing*, vol. 57, Jul. 2023, Art. no. 100752, doi: [10.1016/j.asw.2023.100752](https://doi.org/10.1016/j.asw.2023.100752).
- [72] N. Tenhundfeld, "Two birds with one stone: Writing a paper entitled 'ChatGPT as a tool for studying human-AI interaction in the wild' with ChatGPT," Feb. 2023. [Online]. Available: https://www.researchgate.net/publication/368300523_Two_Birds_With_One_Stone_Writing_a_Paper_Entitled_ChatGPT_as_a_Tool_for_Studying_Human-AI_Interaction_in_the_Wild_with_ChatGPT
- [73] R. Parasuraman and C. Wickens, "Humans: Still vital after all these years of automation," *Hum. Factors*, vol. 50, no. 3, pp. 511–520, Jun. 2008, doi: [10.1518/001872008X312198](https://doi.org/10.1518/001872008X312198).
- [74] T. O'Neill, N. McNeese, A. Barron, and B. Schelble, "Human-autonomy teaming: A review and analysis of the empirical literature," *Hum. Factors*, vol. 65, no. 5, pp. 904–938, Oct. 2020, doi: [10.1177/0018720820960865](https://doi.org/10.1177/0018720820960865).
- [75] J. R. Hayes, "Modeling and remodeling writing," *Written Commun.*, vol. 29, no. 3, pp. 369–288, Jul. 2012, doi: [10.1177/0741088312451260](https://doi.org/10.1177/0741088312451260).
- [76] J. Wilson, N. G. Olinghouse, and G. N. Andrada, "Does automated feedback improve writing quality?," *Learn. Disabilities*, vol. 12, no. 1, pp. 93–118, 2014.
- [77] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Trans. Syst., Man, Cybern., A, Syst. Humans*, vol. 30, no. 3, pp. 286–297, May 2000, doi: [10.1109/3468.844354](https://doi.org/10.1109/3468.844354).
- [78] L. Onnasch, C. D. Wickens, H. Li, and D. Manzey, "Human performance consequences of stages and levels of automation: An integrated meta-analysis," *Hum. Factors*, vol. 56, no. 3, pp. 476–488, 2014, doi: [10.1177/0018720813501549](https://doi.org/10.1177/0018720813501549).
- [79] D. B. Kaber, "Issues in human-automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation," *J. Cogn. Eng. Decis. Mak.*, vol. 12, no. 1, pp. 7–24, 2018, doi: [10.1177/1555343417737203](https://doi.org/10.1177/1555343417737203).
- [80] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004, doi: [10.1518/hfes.46.1.50_30392](https://doi.org/10.1518/hfes.46.1.50_30392).
- [81] R. C. Mayer and J. H. Davis, "The effect of the performance appraisal system on trust for management: A field quasi-experiment," *J. Appl. Psychol.*, vol. 84, no. 1, 1999, Art. no. 123, doi: [10.1037/0021-9010.84.1.123](https://doi.org/10.1037/0021-9010.84.1.123).
- [82] C. C. Tossell, P. Kortum, C. Shepard, A. Rahmati, and L. Zhong, "You can lead a horse to water but you cannot make him learn: Smartphone use in higher education," *Brit. J. Educ. Technol.*, vol. 46, no. 4, pp. 713–724, Jun. 2014, doi: [10.1111/bjet.12176](https://doi.org/10.1111/bjet.12176).
- [83] M. Donnermann, P. Schaper, and B. Lugin, "Integrating a social robot in higher education—A field study," in *Proc. IEEE 29th Int. Conf. Robot Hum. Interact. Commun.*, 2020, pp. 573–579.
- [84] A. Roth, S. Ogrin, and B. Schmitz, "Assessing self-regulated learning in higher education: A systematic literature review of self-report instruments," *Educ. Assessment, Eval. Accountability*, vol. 28, pp. 225–250, 2016, doi: [10.1007/s11092-015-9229-2](https://doi.org/10.1007/s11092-015-9229-2).
- [85] I. Amaro, P. Barra, A. Della Greca, R. Francese, and C. Tucci, "Believe in artificial intelligence? A user study on the ChatGPT's fake information impact," *IEEE Trans. Comput. Social Syst.*, early access, doi: [10.1109/TCSS.2023.3291539](https://doi.org/10.1109/TCSS.2023.3291539).
- [86] A. Choudhury and H. Shamszare, "Investigating the impact of user trust on the adoption and use of ChatGPT: Survey analysis," *J. Med. Internet Res.*, vol. 25, 2023, Art. no. e47184.
- [87] S. Noy and W. Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, vol. 381, no. 6654, pp. 187–192, 2023.