

# Teacher-Scholar: The Journal of the State Comprehensive University

---

Volume 13 | Issue 1

Article 1

---

January 2024

## Are State Comprehensive Institutions Prepared for AI? Not yet.

Brent J. Goertzen

*Fort Hays State University*, [bjgoertzen@fhsu.edu](mailto:bjgoertzen@fhsu.edu)

Brett L. Whitaker

Donnette J. Noble

Justin Greenleaf

Ryan D. Olsen

Follow this and additional works at: <https://scholars.fhsu.edu/ts>



Part of the [Curriculum and Instruction Commons](#), [Educational Methods Commons](#), [Higher Education Commons](#), and the [Leadership Studies Commons](#)

---

### Recommended Citation

Goertzen, Brent J.; Brett L. Whitaker; Donnette J. Noble; Justin Greenleaf; and Ryan D. Olsen (2024) "Are State Comprehensive Institutions Prepared for AI? Not yet.," *Teacher-Scholar: The Journal of the State Comprehensive University*. Vol. 13: Iss. 1, Article 1.

DOI: 10.58809/CXOM8451

Available at: <https://scholars.fhsu.edu/ts/vol13/iss1/1>

This Article is brought to you for free and open access by the Peer-Reviewed Journals at FHSU Scholars Repository. It has been accepted for inclusion in *Teacher-Scholar: The Journal of the State Comprehensive University* by an authorized editor of FHSU Scholars Repository. For more information, please contact [ScholarsRepository@fhsu.edu](mailto:ScholarsRepository@fhsu.edu).

## **Are State Comprehensive Institutions Prepared for AI? Not Yet**

**Brent J. Goertzen**

Fort Hays State University

**Brett L. Whitaker**

Fort Hays State University

**Donnette J. Noble**

Fort Hays State University

**Justin Greenleaf**

Fort Hays State University

**Ryan D. Olsen**

Fort Hays State University

The rise of artificial intelligence (AI) in higher education has profound implications for the way in which faculty teach and students learn. As educators strive to understand the ways in which AI technologies might change the future of higher education, the breadth of research on this topic is concerning. This study illustrates some of the challenges and implications of emerging AI technologies by demonstrating potential challenges and discussing their implications.

There are a myriad of factors that disrupt teaching and learning—illness, natural disasters, gun violence, adverse weather, technological outages, loss of talent or skills, etc. These disturbances differ in spatial scale (e.g., local, regional, national, or globally) and often differ in temporal scale by spanning a day, a week, a semester, a year, or even more. The loci of disruptions also vary by focusing on the individual, a course, a program, or institution (Blanford, et al., 2022). Anyone teaching or learning in March 2020 will likely remember how educators worldwide suddenly pivoted to online learning due to the COVID-19 pandemic—a period of time DeVaney and Quintana (2020) called the “month of whack-a-mole” (para. 5). Whack-a-mole is the arcade game where people use a mallet to bop moles on the head as they randomly appear out of holes. The term has become synonymous with situations wherein repeated efforts to solve a particular problem are exacerbated by the problem continuing to bubble up in other spaces (Macmillan, 2023). That particular moment in time created space for an “essential step on the road to future preparedness” (DeVaney & Quintana, 2020, para. 6) and furthered the case for resilient teaching and adaptive leadership. The authors remind us, however, that we shouldn’t “give up on the present in order to plan for the future, nor should [we] look ahead to the future before solving the problems [currently] at our doorstep” (para. 1).

With its intellectual, emotional, and service-providing nature, teaching is complex and challenging (Wang, 2021). Artificial intelligence (AI) and other

stressors are not going away. Thus, planning for the future educators must engage in resilient teaching. Defined as a “multifaceted, dynamic, process comprised of the interaction of personal and contextual resources” (p. 2), resiliency – the ability to bounce back from stressors – allows educators to adapt to their circumstances and generate positive outcomes amid unusual circumstances. A growing body of research signifies the necessity of integrating teacher resilience into educational systems to assist educators in coping with their realities; however, few nations have taken the necessary measures to incorporate operational steps in terms of a systemic application (Wang, 2021).

Developing a systemic application for AI, where the technological changes are fast and furious, seems like a monumental task. As educators we must navigate the uncertain terrain of AI while honoring institutional values and not losing sight of the north star that guides our work while we diligently shape the future (DeVaney & Quintana, 2020). One thing is certain, we cannot frame teaching resiliency around bouncing back, we must frame it around “bouncing forward” (Blanford, et al., 2022, p. 16) and shape opportunities for the future through adaptive leadership in the AI arena.

Adaptive leadership is an emergent process wherein people with varied backgrounds, talents, skills, beliefs, and values interact to solve problems (Yukl, 2013). Heifetz (1994) describes adaptive leadership as strategies and practices that people create to facilitate change when confronted with complexities. Adaptive work is messy but is necessary to build capacity to thrive anew after periods of disequilibrium or when faced with challenges (Heifetz et al., 2009) and “the inclusion of competing value perspectives may be essential to adaptive success” (Heifetz, 1994, p. 23). Immersed in an increasingly AI driven world, it is imperative that we learn to navigate the related opportunities and threats (Iansiti & Lakhani, 2020) to protect the integrity of our work and that of our students.

In an effort to explore the adaptive challenges associated with this new era of AI, the authors developed this article, the purpose of which is three-fold: 1) to conduct an exercise among departmental faculty members to examine how effective they were at detecting AI-generated content from human generated content; 2) to describe the outcomes of a departmental brown-bag conversation to debrief the results; and 3) to offer short-term and long-term implications of AI in the context of SCUs.

## Literature Review

**ChatGPT and Large Language Model AIs.** Humans have been pursuing intelligence in machines from the earliest days of computing, and much of the focus on intelligence has centered around the ability of machines to exhibit human-like characteristics. Alan Turing, the pioneering mathematician and computer scientist, proposed the famous “Imitation Game” (later known as the Turing Test) thought

experiment in 1950 as a test of a machine's ability to exhibit intelligent behavior (Moor, 2001). This test centered around the concept that computers that could reliably imitate humans to a degree that other humans could not detect it was a machine were said to be intelligent. Of course, this was merely a thought experiment, and was not intended to be a direct measure of "intelligence", but it does make clear that researchers have strongly associated the use of language as a core element of intelligence from the early days of Artificial Intelligence (AI) development (Collins et al., 2021).

AI as a field has evolved and taken many approaches including modeling human problem solving, formal logic, and increasingly complex search capabilities of large data sets, and efforts to accurately portray animal behavior (Raaijmakers et al., 2023). These efforts have culminated with the most recent generation of AI known as machine learning. Machine learning has dominated the last 20 years of AI work, as it offers new capabilities that were previously not possible. Machine learning, in extremely simple terms, is the process of allowing a computer to use statistical algorithms, based on a large set of training data, to make predictions and/or decisions that it wasn't explicitly programmed to do (Meng et al., 2023). Machine learning, coupled with another new technological innovation known as Large Language Models (LLMs), created an explosion in AI capabilities in the field of natural language modeling beginning around 2018 (Collins et al., 2021).

One particular LLM, known as ChatGPT, developed and published by OpenAI and released to the public as a chatbot on November 30th, 2022, received significant notoriety and coverage. This is not because the tool itself was distinctly different than many of the other LLMs being researched and developed at the time, but rather because this was the first free and widely available, public use LLM that most people and institutions outside of the AI research sphere had encountered (Pocock, 2023). The ChatGPT chatbot was designed to mimic a human conversationalist, through the use of machine learning and statistical modeling of how humans speak. Essentially, the chatbot was trained by reading billions of pieces of writing (known as tokens) on the internet and developed an algorithm to make predictions of what words should follow others to create compelling human language; feedback from humans improved its work and modeling (Ramponi, 2022).

LLMs like ChatGPT have many uses and applications. Some examples are activities such as text mining and analysis of large quantities of written content; summarizing topics and gathering basic information on a topic; serving as an efficiency tool for generating writing on a variety of topics; automating feedback and engagement processes; and assisting with translation, coding, and other language-based tasks. Furthermore, and specific to higher education, the strengths of these AI programs lend themselves to wide application within many traditional forms of assessment, skill demonstration, academic writing, and evaluation. Simply

put, much of what is often asked of students in collegiate courses takes the form of writing, and the production of convincing-sounding writing is what these tools do well.

Despite the capability of ChatGPT and other LLMs to generate human-sounding writing, they have significant limitations, and given the nature of how they work, they are often hard to discern or control. In particular, the issues of hallucination, bias, and academic dishonesty are extremely important (Petkauskas, 2023). Hallucination is a term that AI researchers coined to describe the tendency for LLMs to produce plausible sounding, but factually incorrect or misleading writing (Schlegl & Gardhouse, 2023). This writing can be quite convincing given the high quality of the text and human sounding tone, and likely misleads readers who do not engage in rigorous critical examination of the text, or as Talboy and Fuller (2023) stated, “the inherent illusion of authority and credibility activates our natural tendency to reduce cognitive effort ... making it more likely that outputs will be accepted at face value” (p. 7). This is consistent with previous research into other forms of manipulation showing that impressive language increases the believability of otherwise vacuous assertions (Pennycook et al., 2015).

A second serious limitation of ChatGPT and other LLMs is that they can demonstrate significant bias depending on the data they were trained on and the nature of their programming. Since a significant portion of the data that ChatGPT was trained on was writing on the internet, the tool itself can manifest some of the biases present in that writing (Nkonde, 2023). While the internet contains vast swaths of relevant and reliable information, it also includes things such as hate speech, harmful stereotypes, instigation of violence of all kinds, a general overrepresentation of specific viewpoints, and underrepresentation of those that are more marginalized from an online presence (Voorhis, 2023). While the developers of ChatGPT have taken significant steps to limit the presence of bias and put sufficient guardrails in place to prevent it from producing writing that demonstrates biases, it is far from perfect. The tool itself is value neutral, but the human writing it learned from is not.

ChatGPT is proficient at the core task of generating realistic, simulated human conversation, but upon release to the public, users began exploring the other capabilities of the program, as well as using the chatbot for purposes well beyond casual conversation (Basile et al., 2022). This included some users attempting to “jailbreak” the program; essentially trying to work around the guardrails in place to force the program to produce offensive or biased content (Martindale, 2023). The implications of these applications took OpenAI, as well as many other organizations and institutions, by surprise.

Therefore, there are significant ethical, social, and practical concerns about the latest generation of natural language AIs (Lund, 2023). Given the significant implication for academic dishonesty and intellectual property issues more generally

(Ceres, 2023), higher education was one of the industries most vocal about critical implications of this tool. Students may employ ChatGPT to produce all sorts of academic writing, and the detection of the use of these tools is incredibly difficult. As a technology tool, the use of ChatGPT is not inherently inappropriate (we encourage the utilization of many tools to improve writing, such as spell-checker), but submitting materials entirely generated by AI as one's own work is fraught with concern. Furthermore, since ChatGPT generates new writing (as opposed to copying existing writing), it occupies a space that has not previously existed, compared to more familiar forms of academic dishonesty like plagiarism. There may be a dearth of institutional policy and guidance about how to engage with this type of cheating, as well as limited utility from other tools used to control for plagiarism, such as detection software.

**Pedagogical Approaches.** When considering AIs impact on leadership education, and more broadly, other related social science disciplines, it makes sense to think about the instructional approaches used by leadership educators. Fortunately, the question of how to teach leadership has been a significant focus of study in recent years. Even though academic-based leadership programs vary widely in terms of purpose and mission (Brungardt, et al., 2006; Greenleaf, et al., 2018), evidence suggests that the teaching of leadership shares common practices. Common sources of leadership development include classroom-based training, e-learning, group and individual reflection, and developmental assignments (Allen & Hartman, 2008). Exploring these sources provided a foundation for further analysis of specific teaching techniques. Jenkins (2012, 2013), examined often-used teaching strategies employed in undergraduate leadership education and reported that case studies, class discussions, and in-class short writing assignments were common pedagogical techniques of leadership educators in face-to-face courses, whereas discussion boards (both instructor-led and student-led), reflective journals or blogs, and online case studies were commonly used in online leadership instruction (Jenkins, 2016).

These instructional strategies point toward teaching and learning that emphasizes students' ability to generate written work and to critically analyze and evaluate content. These pedagogical strategies were consistent in graduate education. Jenkins (2018) reported, when comparing undergraduate-level to graduate-level instruction, educators at the graduate-level more frequently integrated the use of writing projects, term papers, research projects and leadership development plans.

While the studies cited here acutely focus on field of leadership development, the pedagogical strategies are comparable to other social and behavioral sciences as well as many other academic programs of study within SCUs as we share core purposes of education civic leaders and engaged global citizens. It is common for our associations (e.g., American Association of State Colleges and Universities,

AASCU) to promote initiatives that incorporate service-learning (Shapiro & Mills, 2012) and other pedagogies that are experiential, facilitative, and collaborative (Summit, 2013).

Educators draw upon a wide range of pedagogical approaches to facilitate leadership learning. Yet, among the biggest challenges of teaching and learning is documenting the learning that takes place. Educators must evaluate student responses to a given assignment and corresponding assessment instrument that represents students' knowledge of a content area of interest. Learning is a process internal to an individual. While educators draw upon a wide array of pedagogical methods, we frequently require students to generate written work to express their individual reasoning process of their learning. LLM tools, such as ChatGPT, can be useful for students to express their thinking and learning, but questions remain regarding to the extent of which the written work accurately represents students' own internal reasoning. This is just one of the many ways in which generative-AI will impact teaching and learning.

Ultimately, there is nothing inherently wrong with these approaches. However, when AI platforms are introduced to the learning environment and they can easily create usable content, issues of academic integrity become a serious concern.

## Methods

**Generating Responses from a Student Assignment.** The researchers collected student responses based on course content whereby many faculty and adjuncts of the department would be familiar. StrengthsQuest (Clifton & Anderson, 2001) is a required activity in a course on team dynamics whereby students learn about their personal and team members' strengths and determine pathways to harness their strengths for team effectiveness. The researchers selected a short-answer, essay question that invites students to both, identify their top two strengths and express how those strengths will be particularly useful in the team context throughout the semester.

One of the researchers interacted with ChatGPT to create the AI-generated responses. The initial prompt given to ChatGPT was "write a 30-80 word response to an exam question that states: 'select two strengths (referring to StrengthsQuest or Clifton), and describe how each might help a person work in a team'. Be specific and write in first-person." A secondary prompt specified, "select different strengths and a different example. Change up the writing a little." This was repeated until ChatGPT generated nine responses. The responses were recorded verbatim as produced by ChatGPT, with no editing of any kind by the researchers.

Another researcher recently taught multiple sections of the team dynamics class and assembled 42 student responses. The researcher removed outliers, either exceptionally long or very short responses, resulting in 27 student responses. Next,

a random number generator was used to further narrow down the student responses to nine.

The researchers combined the nine human and nine AI-generated responses and determined the combined items may be overly onerous for participants. Believing twelve items would provide meaningful results, the researchers again engaged a random number generator to randomly select twelve of the responses to be included in the final survey resulting in six human and six AI-generated responses.

**Sample and Survey Design.** Participants for the study were selected from the full-time and adjunct faculty affiliated with an academic leadership studies program part of an SCU located in the US Midwest. The total population was 18 (seven full-time faculty and 11 adjunct faculty). Seven completed responses were included for analysis representing four full-time faculty and three adjunct faculty.

Participants were invited by email to interact with the online survey. It was explained that the purpose of the study was to evaluate how effective faculty are at detecting the difference between human or AI-generated content. Participants were instructed to review 12 “student assignments” representing a short answer, essay question and then complete the following tasks: 1) evaluate the assignment on three dimensions of the grading rubric (Understanding of Concepts; Application of the Concepts; and Writing Quality); 2) assign a point value earned for the response (0 to 100); 3) reply to an item indicating whether you believe the assignment was either human or AI-generated; and 4) respond to an open-ended question providing a brief rationale as to why they believed the assignment was either AI or human-generated.

## Results

**Grade and Rubric Evaluation.** Results regarding the grades assigned the response are reported in Table 1. The overall average score assigned to the responses was 87.67 (out of 100 points) representing a percent grade. The responses identified as AI-generated were 87.05 whereas the responses identified as human-generated scored slightly higher at 87.30. Put differently, what participants perceived to be human responses and AI-generated responses, on average, they graded very similarly. However, the assigned grades to the responses that were actually generated by AI (90.79) were substantially higher than the grades assigned to human responses (84.55). The difference of six-points is analogous to ‘half a letter-grade’ higher scores for AI-generated responses compared to those actually generated by humans.



**Table 1**  
*Average Grade Assigned to Responses*

	<b>Score (out of 100)</b>
<b>Overall average</b>	<b>87.67</b>
<b>Perceived</b>	
AI (e.g., ChatGPT)	87.05
Human	87.30
<b>Actual</b>	
AI (e.g., ChatGPT)	90.79
Human	84.55

A similar pattern emerged in the results from examining rubric ratings (see Table 2). Regarding the rubric ratings on the Understanding of Concepts dimension, what participants perceived as AI responses scored on average 2.78 (on a four-point scale), whereas what participants perceived as human generated responses scored nominally higher at 2.82. However, the responses actually generated by AI scored substantially higher (3.10) compared to the human generated responses (3.67). Evaluation of the Application of Concepts yielded a similar pattern of results. The responses perceived as AI-generated scored 2.71 which was slightly lower than responses perceived as human-generated, 2.79. Again, the responses actually created by AI were substantially higher (3.02) than human-generated responses (2.57). Finally, analysis on the Writing Quality dimension of the rubric revealed the responses perceived to be AI-generated (3.29) scored higher than the responses perceived to be human-generated (2.89). However, the spread was more pronounced when examining the scores assigned to the responses actually generated by AI (3.48) compared to the responses actually generated by humans (2.67).

**Table 2**  
*Average Evaluation of Rubric Dimensions to Responses*

<b>Rubric Dimensions</b>	<b>Average</b>
<b>Understanding of Concepts</b>	<b>2.88</b>
<b>Perceived</b>	
AI (e.g., ChatGPT)	2.78
Human	2.82
<b>Actual</b>	
AI (e.g., ChatGPT)	3.10
Human	2.67
<b>Application of Concepts</b>	<b>2.77</b>
<b>Perceived</b>	
AI (e.g., ChatGPT)	2.71
Human	2.79
<b>Actual</b>	
AI (e.g., ChatGPT)	3.02
Human	2.57
<b>Writing Quality</b>	<b>3.07</b>
<b>Perceived</b>	
AI (e.g., ChatGPT)	3.29
Human	2.89
<b>Actual</b>	
AI (e.g., ChatGPT)	3.48
Human	2.67

**Accuracy and Rationale in Determining AI or Human Generated Responses.** Overall, participants were able to correctly identify responses as AI or human-generated at 70.2% (see Table 3). A more detailed examination of accuracy indicated that participants were more able to detect when a response was generated by humans (78.6%) whereas they were only able to correctly identify AI-generated responses 61.9% of the time. A further comparison was conducted between full-time and adjunct faculty. Where full-time faculty were able to correctly identify human or AI-generated responses 83.3 percent of the time, adjunct faculty only were able to correctly identify slightly more than half of the responses (52.78%).

**Table 3**

*Percent Accuracy of Identifying Responses as AI or Human Generated*

	<b>Accuracy (%)</b>
<b>Accuracy (overall)</b>	<b>70.2</b>
AI (e.g., ChatGPT)	61.9
Human	78.6
Full-time faculty	83.33
Adjunct faculty	52.78

Qualitative comments provided by participants as rationale as to why they believed responses were either generated by AI or humans were categorized prior to analysis into one of four cells: 1) correctly identified as AI; 2) incorrectly identified as AI; 3) correctly identified as human or 4) incorrectly identified as human. A content analysis was employed to identify themes common to each cell and theme frequencies (see Table 4). When participants correctly identified a response as AI-generated, they commonly cited the response as using “almost perfect” or “sterile language” as well as occasionally mis-identifying and defining key concepts. For example, one participant stated, “the response was very generic and did not provide specific examples. It read like a textbook and was almost perfect in its language.”

When participants incorrectly identified AI-generated responses they often cited the “use of first-person pronouns” and “affective domain” as reasons they believed humans created the response. For instance, one participant commented the response “uses the first person ‘I am confident.’ The phrase ‘better understand and relate to my teammates’ seems more contextual to a class project.” Another participant, when incorrectly identifying an AI-generated response as human said, “I do not believe the AI can apply appropriate Affective Domain levels within answers.”

**Table 4***Themes of Rationale When Identifying Responses as either AI or Human*

<b>Correctly Identified as AI (frequency)</b>	<b>Incorrectly Identified as AI (frequency)</b>
<ul style="list-style-type: none"> <li>• Perfect/Sterile/Formulaic Language (13)</li> <li>• Unusual Use of Language, Buzz Words (not a student voice) (6)</li> <li>• Inaccurate use of Content (4)</li> <li>• Lack of Personal Example (2)</li> <li>• Lack of affective domain (1)</li> </ul>	<ul style="list-style-type: none"> <li>• First person (e.g., I) (5)</li> <li>• Affective domain (3)</li> <li>• Word choice/ Human descriptors (e.g., bigger picture) (3)</li> <li>• Tone (2)</li> </ul>
<b>Correctly Identified as Human (frequency)</b>	<b>Incorrectly Identified as Human (frequency)</b>
<ul style="list-style-type: none"> <li>• Poor writing/grammar (17)</li> <li>• Informal/conversational writing (5)</li> <li>• Personalized writing (4)</li> <li>• Misinterpreting the Concepts (3)</li> <li>• Specific use of Concepts (2)</li> <li>• Use specific reference &amp; page number (2)</li> <li>• Formatting (e.g., bullet points) (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Formulaic writing/Organization (3)</li> <li>• Insufficient affective domain (2)</li> <li>• Lacks personalization (1)</li> <li>• Too many “I” statements (1)</li> <li>• Natural grammar (1)</li> </ul>

Participants, when correctly identifying responses as generated by humans, commonly cited writing conventions and quality as key indicators of human created content. Some of the rationale includes “sentence structure gave me the impression this was a student’s writing,” or “rough narrative flow.” Participants also observed the use of specific citations as rationale for human content by expressing comments such as the response “reference directly from the book and edition we provided for the readings” in the course.

Finally, when participants incorrectly identified responses as human-generated content, they likewise commented about the writing style and organization of the response and insufficient affective domain as reasons. For instance, a participant explained “the repetition of ‘having strength in’ seems formulaic. There is first person at the beginning, but it's not really weaved in and goes more to description. The organization of the paragraph also makes me think AI.” Another participant mentioned “I am not seeing sufficient Affective Domain key words” and hence the content must have been AI-generated.

## **Discussion of Outcomes and Implications of Application**

**Recommendations.** The most direct implication of this study became obvious to us at the request of the participants. Immediately upon conclusion of the data collection, participants began asking to learn about which of the statements they viewed were AI generated and also how accurate they were in their judgements and grading assessments. Given these requests and to facilitate professional development among our faculty group, we proceeded to arrange a brownbag session focused on revealing the results of the study and discussing implications for our program.

Faculty participants who attended the brownbag session were stunned. When presented with the preliminary results that have been reported in this manuscript, they expressed a great deal of surprise at both how convincingly the ChatGPT tool had produced the homework, as well as how poor they were at detecting the characteristics of human or AI generated writing. In particular, faculty members conveyed that they believed that most of the assignments we regularly utilize, and especially reflective writing, would be very resilient against this type of artificially generated content, and they were shocked to see how that was not the case. The hubris with which they approached this study, and the subsequent humbling in the face of their own capabilities was incredibly valuable as a growth exercise, and we hope that other readers of this manuscript will have a similar reaction.

We invite you to engage with a small interactive element that will hopefully provide you with a similar revelatory experience (see Interactive Reader Activity). Included in the activity are three samples that were utilized in the study. Each of these was either human written or AI generated. We encourage you to take a moment and review these samples, making judgements about which you believe are generated and which are authentic human writing. After you are finished, you may view the answers at the end of the activity.

The implications of natural language AIs are significant, pervasive, and not yet well understood. As we consider outcomes and implications, educators ought to contemplate both short-term controlling strategies, and as well as long-term adaptation strategies. It is tempting to initially see natural language AI tools as merely the latest iteration of technology tools that students are using to cheat their way past writing and engaging with coursework. However, these tools are powerful new ways of engaging with information sharing, and they are unlikely to go away or be irrelevant outside of an academic context. Therefore, the authors encourage readers to be reflective in considering how their own pedagogical or andragogical approaches to instruction, course and program outcomes, and learning assessment should evolve to embrace these tools in the long term. While educators are adapting to the long-term impact of natural language AIs, they may also wish to control for the potential of students to cheat in their courses. In this section, we delineate

several strategies at both the short- and long-term orientation and encourage readers to consider these strategies as may be appropriate for their own programs.

**Short-Term Implications.** There are immediate concerns regarding academic integrity relative to students' use of these AI tools. Many faculty members may wish to deploy controlling strategies in the coming semesters to help ensure the rigor of their courses. Here are a few methods to do so.

First, it is important to clearly communicate to students what is and is not appropriate in your own course(s). Faculty should consider including a statement in their syllabus outlining the course policies regarding use of AI language generation, appropriate use cases, and consequences for misuse of the tools. Faculty may also consider engaging with students in examining the strengths and weaknesses of ChatGPT and similar tools through in-class exercises, both to help educate students about the limitations of these tools, as well as to better understand how they can make work more efficient.

For faculty wishing to prohibit or severely limit the use of ChatGPT or other LLMs amongst their students, the best way to do so is through assessment design. In particular, faculty should consider limiting low-order writing assessments that are especially susceptible to algorithmic writing. Assignments such as defining concepts, describing the basic components of a theory, or merely listing, restating or otherwise summarizing information are exactly the kind of writing that these tools do best, and therefore are at the greatest degree of susceptibility for their use. This is not to say that low-order outcomes are not valuable from an educational standpoint, but rather, faculty should target their learning assessments on higher order outcomes that are built upon low order knowledge base. For example, a faculty member teaching adaptive leadership might consider using an objective knowledge test for outcomes connected to remembering and understanding, and then target a writing project assessment for outcomes focused on evaluation or creation.

However, even with the best design, there is always the possibility of students submitting algorithmically generated writing as their own work, and educators will need to consider ways to check for this as part of learning assessment. Some parallels can be drawn from techniques already deployed to detect plagiarism, such as asking students to present their written work in an oral speech or conversation, using in-class or handwritten writing, and utilizing group writing assessments. Another strategy to detect plagiarism is by using AI detection software to detect AI generated writing. While there are some promising tools becoming available, these tools are not fool proof, and may often produce false-positive or false-negative results. Furthermore, these new tools may not have an enforceable standing within an institutional setting, unlike more established protocols and tools such as plagiarism detection software (Drozdowski, 2023). Furthermore, unlike traditional plagiarism, since there is no source writing to compare to, even if the software flags

something as likely being AI generated, there may be no definitive way to prove that a student did not actually produce the writing.

**Long-Term Implications.** There are significant, far-reaching implications associated with the introduction of LLM technology that will likely revolutionize how we ‘do leadership’ along with the teaching and learning of leadership. Specifically, three general implications include: 1) how will humans interact to address complex challenges facing our organizations and society?; 2) how will educators engage in the teaching and learning of leadership?; and 3) how will the larger academy and our knowledge producing systems adapt to keep up with the unremitting technology revolution?

Leadership, being a human-oriented enterprise, is widely understood as a social process where humans interact to identify complex challenges and work to achieve mutual purposes (Burns, 1979; Heifetz et al., 2009). This continued need for human interaction will not change; however, the first key implication of LLMs and AI technology is how they will compel us to reconsider how we interact. AI tools have already been instrumental in analyzing “big data” to identify trends and forecasting future opportunities and thereby identifying adaptive challenges and potential solutions. LLMs may be able to perform various other technical tasks normally accomplished by humans such as the writing of a strategic plan. Certainly, human effort is required to input parameters into the AI tool, but they could be freed from the process of physically writing the strategic plan or other technical tasks to dedicate more energy toward the execution of other human-to-human related activities.

Second, there are significant implications regarding teaching and learning. As demonstrated in the present study, LLMs can produce content that readily and reasonably accurately described key concepts related to course content. Reimagining assignments and learning assessments is necessary in the era of LLMs. One strategy a faculty recently deployed was to create a detailed writing assignment prompt for his students and directed students to then enter the assignment prompt into ChatGPT (Howell, 2023). Then, applying predetermined grading criteria, he required students to provide written feedback, as well as to critically examine the sources cited by the AI tool. Not only did this learning exercise enhance students’ understanding of course content but also exposed them to the limitations as ChatGPT ‘hallucinated’ some content it produced. By applying this or similar ideas, educators can create assignments that foster greater creative and critical thinking in addition to developing an understanding of foundational concepts associated with theory and practice. This is particularly important given the “decline in people’s ability to think deeply and reflectively in the past few years” (Well, 2023, para. 2).

There are inherent ethical implications regarding the use of LLMs in teaching and learning, not the least of which is related to academic integrity. It is incumbent

upon local institutions, academic units, and individual instructors to partner in concert in determining expectations regarding appropriate or permissible use of AI during the learning experience. We envision a continuum of various options on the use of AI; from its ‘use prohibited’; to ‘use only with prior permission’; to ‘use only with acknowledgement.’ Related aspects to the implications on teaching and learning involve cultural contexts. Institutions of higher education, particularly in many SCUs, are increasingly creating cross-border or global partnerships. Not all cultures hold the same cultural mores and must proactively work to navigate cultural differences to maintain robust and healthy cross-border partnerships.

Nonetheless, one of our primary roles as educators is to prepare our students for the contemporary work environment. The modern workplace desires a productive workforce, among other things, in terms of efficiency and effectiveness. For many, they will perceive LLMs as an important tool to enhance employee productivity. It is our responsibility to prepare today’s learners to effectively engage in the modern workplace. As such, prohibiting the use of AI in our courses is not necessarily a wise option. We must assist our learners as to how to use AI tools and use them well.

A third implication relates to the broader academy and conventional knowledge producing systems. Knowledge perceived as credible is that which has been rigorously vetted through a process of review. Blind, peer-review publications are often considered the ‘gold standard’ of knowledge production however, the blind, peer-review process can be laboriously slow and inefficient. In some circumstances, the review process for academic journals may take well over a full calendar year from the initial submission to final publication (if accepted). While the Teacher-Scholar journal has a reputation for being expedient in terms of the review and publication process, we acknowledge that since the inception of our research idea when the first public-facing versions of generative-AI tools (e.g., ChatGPT) became available to the initial submission of our manuscript, ChatGPT unveiled two new generational updates, not to mention the many other generative-AI tools that have been launched by its competitors. Our point? Conventional knowledge producing systems are currently underprepared to keep pace with this advancing technology.

**Limitations.** The researchers of the application brief acknowledge several important limitations. For instance, the research design represented a small set of faculty from a single academic program. The purpose of the study was not to produce generalizable results, rather it was to raise faculty awareness about AI and potential challenges as they engage students in their courses.

Another limitation is that the researchers took verbatim responses of the text generated from the AI prompt which under-represents the potential of what it can produce. Any savvy student could readily modify the text to more accurately fit an



assignment prompt in our classes as well as have it be more representative of the quality of the student's work.

Researchers of the present study are not computer scientists or experts in the field of AI. At best, we consider ourselves novices. Yet, we share a deep interest in AI as a tool useful to enhancing how humans interact and the potential implications of it on the teaching and learning of leadership.

The final limitation relates to the particular point in time the study was conducted. We produced this body of work during the embryonic stages of the public-facing, LLM revolution and the subsequent broad awareness of how it was beginning to impact higher education. Since the time of when data collection was completed, ChatGPT has already unveiled two new iterations and undoubtedly has greater capacity to overcome the limitations of previous versions.

## **Conclusion**

Educators engage in a variety of teaching strategies, and the introduction of AI into the educational environment presents an adaptive challenge that has yet to be fully explored within local disciplines and more broadly at our SCUs. The examples of ChatGPT and LLMs described in this study demonstrate how work generated by AI tools can be difficult to identify and prove as being algorithmically generated. These evolving technologies are forcing educators to reconsider how they will interact with students as well as redefine what will constitute scholarly work in the future.

The value of education and how it should be pursued will need to be revisited as AI becomes more pervasive (Klutka et. al., 2020). Educational institutions are faced with an array of decisions when it comes to AI. For example, will they be among the early adopters and build the institutional memory within their operations now and reap the early rewards or wait until the applications are more diverse and scalable? The first dedicated degree in AI was launched at Carnegie Mellon University in 2018, and they are already reaping the aforementioned early rewards. Even so, "Higher education as an industry may be late to AI in comparison to the corporate sector, but we can extrapolate from how it plays out for businesses to forecast for our landscape" (p. 29). Discrete skills will likely be among the first to become automated while the more complex skills remain within the human scope (Klutka, et. al., 2020)—but for how long? That is precisely why educators must be proactive when it comes to mitigating the hazards and enjoying the advantages of AI.

## **Interactive Reader Activity**

**Sample Submissions.** The following are examples of three responses that were provided to participants within the study. We invite you to engage with this work by reading these and responding to two basic questions for each. First, try to

determine whether this response is AI generated or human written. Following that, briefly record the central reason or two why you feel this way. The correct answers can be found at the end of this activity, but we encourage you to record your responses before looking at the answer key for maximum impact.

**Sample #1.** “Drawing from my strengths assessment, two strengths that I believe would contribute positively to team dynamics are strategic thinking and empathy. With strategic thinking, I can take a big-picture perspective and consider different options and outcomes, enabling the team to make informed and decisive decisions. Furthermore, empathy allows me to better understand and relate to my teammates, creating a supportive and inclusive environment where each person feels valued and heard. By combining these two strengths, I am confident that I can contribute to a collaborative team that is not only effective in achieving its goals but is also compassionate and respectful towards each other.”

AI or Human? \_\_\_\_\_

Why? \_\_\_\_\_

**Sample #2.** “Two of my strengths, achiever and positivity, are incredibly beneficial when working in a team. As an achiever, I am always motivated to meet and exceed goals, which can inspire and drive my team members to do the same. My positivity is contagious, and it enables me to uplift and encourage my colleagues when things get tough. This not only boosts the team's morale but also makes everyone feel valued and appreciated. Ultimately, with my strengths, I can bring productivity and optimism to the team, which in turn, increases our chances of success.”

AI or Human? \_\_\_\_\_

Why? \_\_\_\_\_

**Sample #3.** “Based on my strengths, I believe that positivity and empathy would be valuable qualities to bring to a team. By having a positive attitude, I can maintain a motivating and optimistic environment, helping team members feel encouraged and engaged in the task at hand. Furthermore, through practicing empathy, I can put myself in my teammates' shoes, truly understand their perspectives and emotions, and make decisions that benefit everyone. These

strengths allow me to build stronger relationships, resolve conflicts effectively, and encourage collaboration within the team. In conclusion, having a positive and empathetic mindset can be a huge asset in any team setting.”

AI or Human? \_\_\_\_\_

Why? \_\_\_\_\_

**Interactive Reader Activity: Answer Key.** The three samples presented in the Interactive Reader Activity consist of two AI generated responses and one human written response. Samples 1 and 3 are AI, and sample 2 was written by a student.

If you were able to successfully identify all three, congratulations! However, if you struggled with this, or were incorrect in your identification, your experience is consistent with the participants in the study who had a difficult time accurately assessing these responses.

To conclude this activity, we encourage you to reflect on two points. First, consider how likely it would be that you would detect the use of AI generated content in a regular course setting when you’re not primed to be especially critical. We assume that we would almost always take the students’ work at face value and fail to recognize all but the most flagrant uses of these tools in our current courses. Second, you may take some time and compare your justifications provided in Interactive Reader Activity with those of the participants in the study. Perhaps you had some of the same notions about what makes something sound more “human”, and perhaps those notions are faulty.

Thanks for engaging with this activity. We hope it was illuminating.

### References

- Allen, S. J. & Hartman, N. S. (2008). Leader development: An exploration of sources of learning. *Organization Development Journal*, 26(2), 75-87.
- Basile, V., Caselli, T., Balahur, A., & Ku, L. W. (2022). Bias, subjectivity and perspectives in natural language processing. *Language and Computation*, (5), 926435. <https://doi.org/10.3389/frai.2022.926435>
- Blanford, J. I., Bowlick, F., Gidudu, A., Gould, M., Griffin, A. L., Kar, B., Kemp, K., de Roiste, M., de Sabbata, S., Sinton, D, Strobl, J., Tate, N., Toppen, F., & Unwin, D. (2022). Lockdown lessons: an international conversation on resilient GI science teaching. *Journal of Geography in Higher Education*, 40(1), pp. 7-19.

- Brungardt, C. B., Greenleaf, J., Brungardt, C. J., & Arensdorf, J. (2006). Majoring in leadership: A review of undergraduate leadership degree programs. *Journal of Leadership Education*, 5(1), 4-25.
- Burns, J.M. (1979). *Leadership*. New York: Harper and Row.
- Ceres, P. (2023). *ChatGPT is coming for classrooms*. Wired. <https://www.wired.com/story/chatgpt-is-coming-for-classrooms-dont-panic/>
- Clifton, D.O., & Anderson, E. (2001). *StrengthsQuest: Discover and develop your strengths in academics, career, and beyond*. Gallup Press.
- Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, 60.
- DeVaney, J. & Quintana, R. (2020). *Preparing for future disruption: Hybrid, resilient teaching for a new instructional age*. Inside Higher Ed. <https://www.insidehighered.com/blogs/learning-innovation/preparing-future-disruption-hybrid-resilient-teaching-new-instructional>
- Drozdzowski, M. J. (2023). *Testing Turnitin's new AI detector: How accurate is it?* BestColleges. <https://www.bestcolleges.com/news/analysis/testing-turnitin-new-ai-detector/>
- Greenleaf, J., Kastle, S., Sramek, K., & Brungardt, B. (2018). Majoring in leadership: Issues of consistency and credibility. *Journal of Leadership Studies*, 12(2), 53-55.
- Heifetz, R. (1994). *Leadership without easy answers*. Belknap Press.
- Heifetz, R., Grasnaw, A., & Linsky, M. (2009). *The practice of adaptive leadership: Tools and tactics for changing your organization and the world*. Cambridge Leadership Association.
- Howell, C.W. (June 24, 2023). *ChatGPT and the university's existential crisis*. Writing on Religion, Science, and Technology. <https://www.cwhowell.com/chatgpt-and-the-universitys-existential-crisis/>
- Iansiti, M. & Lakhani, K.R. (2020). *Competing in the age of AI: How machine intelligence changes the rules of business*. Harvard Business Review. <https://hbr.org/2020/01/competing-in-the-age-of-ai>
- Jenkins, D. M. (2012). Exploring signature pedagogies in undergraduate leadership education. *Journal of Leadership Education*, 11(1), 1-27.
- Jenkins, D. M. (2013). Exploring instructional strategies in student leadership development programming. *Journal of Leadership Studies*, 6(4), 48-62.
- Jenkins, D. M. (2016). Teaching leadership online: An exploratory study of instructional and assessment strategy use. *Journal of Leadership Education*, 15(2), 129-149.

- Jenkins, D. M. (2018). Comparing instructional and assessment strategy use in graduate- and undergraduate-level leadership studies: A global study. *Journal of Leadership Education, 17*(1), 75-92.
- Jenkins, D.M. (2020). What the best leadership educators do: A sequential explanatory mixed methods study of instructional and assessment strategy use in leadership education. *Journal of Leadership Education, 19*(4), 37-55.
- Klutka, J., Ackerly, N., & Magda, A.J. (2020). *Artificial intelligence in higher education: Current uses and future applications*. Wiley Educational Services. <https://universityservices.wiley.com/wp-content/uploads/2020/12/201811-AI-in-Higher-Education-TLH-with-new-bage.pdf>
- Lund, J. (2023). *AI is exciting and an ethical minefield: 4 essential reads on the risks and concerns about this technology*. The Conversation. <https://theconversation.com/ai-is-exciting-and-an-ethical-minefield-4-essential-reads-on-the-risks-and-concerns-about-this-technology-204444>
- Martindale, J. (2023). *How to jailbreak ChatGPT: get it to really do what you want*. Digitaltrends. <https://www.digitaltrends.com/computing/how-to-jailbreak-chatgpt/>
- Macmillan (2023). *Whack-a-mole*. Macmillan.com dictionary. <https://www.macmillandictionary.com/us/dictionary/american/whack-a-mole>
- Moor, J. H. (2001). The status and future of the Turing Test. *Minds and Machines, (11)*1, 77-93.
- Nkonde, M. (2023). *ChatGPT: New AI system, old bias?* Mashable. <https://mashable.com/article/chatgpt-ai-racism-bias>
- Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making, 10*(6), 549-563.
- Petkauskas, V. (2023). *ChatGPT's answers could be nothing but a hallucination*. CyberNews. <https://cybernews.com/tech/chatgpts-bard-ai-answers-hallucination/>
- Pocock, K. (2023). *What is ChatGPT and what is it used for?* PCGuide. <https://www.pcguides.com/apps/what-is-chat-gpt/>
- Raaijmakers, S., Cremers, A., Krahmer, E., & Westera, M. (2023). Conversational AI. *Language and Computation, (6)*, 1203910. <https://doi.org/10.3389/frai.2023.1203910>
- Ramponi, M. (2022). *How ChatGPT actually works*. AssemblyAI. <https://www.assemblyai.com/blog/how-chatgpt-actually-works/>
- Schlegl, L. A. & Gardhouse, K. (2023). *Reducing bias in ChatGPT with PrivateGPT*. PrivateAI. <https://www.private-ai.com/2023/06/06/reducing-bias-in-chatgpt-with-privategpt/>

- Shapiro, M. S. and Mills, S. (2012). Educating globally competent citizens: An AASCU Red Balloon Project. *Teacher-Scholar: The Journal of the State Comprehensive University*, 4(1), 55-67. <http://doi.org/10.58809/CTGI7342>
- Summit, J. (2013). Global citizenship demands new approaches to teaching and learning: AASCU's Global Challenges Initiative. *Change: The Magazine of Higher Learning*, 45(6), 51-57. <https://doi.org/10.1080/00091383.2013.842109>
- Talboy, A. N. & Fuller, E. (2023). Challenging the appearance of machine intelligence: Cognitive bias in LLMs. *Arxiv*, 2304.01358.
- Voorhis, S. V. (2023). *ChatGPT: Did big tech set up the world for an AI bias disaster?* Harvard Business School. <https://hbswk.hbs.edu/item/chatgpt-did-big-tech-set-up-the-world-for-ai-bias-disaster>
- Wang, Y. (2021). Building teachers' resilience: practical applications for teacher education in China. *Frontiers in Psychology*, 12, 738606. <https://doi.org/10.3389/fpsyg.2021.738606>
- Well, T. (2023). *The decline of critical thinking skills*. Psychology Today. <https://www.psychologytoday.com/us/blog/the-clarity/202306/the-decline-of-critical-thinking-skills?amp>
- Meng, W., Zhang, X., Ru, B., & Guan, Y. (2023). A machine learning approach to real-world time to treatment discontinuation predication. *Advanced Intelligent Systems*, (5)4, 2200254. <https://doi.org/10.1002/aisy.202200254>
- Yukl, G. (2013). *Leadership in organizations* (8<sup>th</sup> ed). Pearson.