Critical Trust in Generative AI: Suspending Disbelief and Developing Critical Friend Groups Among University Students

Timothy Mattison (https://orcid.org/0000-0002-5292-0760), Elizabeth Wilkins (https://orcid.org/0000-0002-7807-7345), and Julie Conrad, University of Southern Indiana, USA

Abstract

In a 2025 international study involving 11,706 undergraduate students, 29% overall (67% in the U.S.) reported regular use of generative artificial intelligence (GenAI) programs. This could be problematic, since GenAI outputs often contain randomly parroted falsehoods, and students could overuse the technology. This manuscript proposes a new framework, critical trust, to help university students balance trust and mistrust in the validity of information in GenAI outputs. A critical trust framework could also help students make ethical decisions about how and when to use GenAI applications in their academic work. University students can employ critical trust in *critical friend groups*, which could be especially beneficial for students from historically marginalized communities without college-educated mentors, helping them avoid accusations of misuse of GenAI and the loss of credibility that can result from plagiarism. The critical trust framework could also provide education researchers with a pathway to develop and test a scale that measures critical trust, ultimately determining the optimal levels to minimize harms from GenAI hallucinations and inappropriate overuse.

Keywords: generative artificial intelligence, AI, GenAI, trust, critical hope, critical friend groups

Introduction

OpenAI's release of ChatGPT in the fall of 2022 and the over 2,000 generative artificial intelligence (GenAI) programs that followed created new opportunities and challenges for university students. Critical trust refers to a calibrated, reflective stance toward generative AI systems that combines informed confidence in their utility with disciplined skepticism about their limitations, biases, and epistemic risks. Rather than unconditional acceptance or wholesale rejection, critical trust requires users to actively evaluate outputs, verify claims, and remain accountable for interpretive and ethical judgment. For instance, GenAI programs such as Gemini, Grammarly, and Otter.ai can save students time and effort by automating notetaking, transcription, copyediting, research, and even the generation of unique text or images. As a result, a growing number of university students are using these programs in their studies. In 2025, Yonder Consulting conducted an international poll of 11,706 undergraduate students ages 18-21 as part of the annual Chegg Global Student Survey (Chegg, Inc., 2025). Polling results revealed that 29% of respondents reported regularly using GenAI programs to assist them in their studies (Chegg, Inc., 2025). This percentage was over twice as high (67%) among respondents from the United States (Chegg, Inc., 2025).

Critical trust refers to a calibrated, reflective stance toward generative AI systems that combines informed confidence in their utility with disciplined skepticism about their limitations, biases, and epistemic risks. Rather than unconditional acceptance or wholesale rejection, critical trust requires users to actively evaluate outputs, verify claims, and remain accountable for interpretive and ethical judgment.

Students' high rates of GenAI use raise serious concerns, given the pervasiveness of hallucinations in GenAI outputs (Maleki et al., 2024). GenAI hallucinations are randomly parroted falsehoods that could cause students to unknowingly spread misinformation in their

assignments and, if detected by professors, could harm students' grades and academic reputation (Maleki et al., 2024). These hallucinations could be especially harmful to students from historically marginalized communities without college-educated parents to coach them through repairing any possible damage to their grades and reputation from GenAI use. Hallucinations are even more of a challenge in the current environment of digital information overload, where there is a scarcity of expert curators for GenAI outputs. GenAI programs also tempt students to overuse the technology, potentially enabling them to bypass meaningful engagement in crucial parts of the learning process, such as analysis, evaluation, connection-making, and generating new ideas.

However, despite problems with GenAI hallucinations and overuse, avoiding the use of GenAI programs is becoming increasingly difficult for university students due to their future employers' expectations. Many private-sector firms and public agencies are exploring the potential integration of GenAI programs to enhance their employees' efficiency and adjust staffing models (Sira, 2023). For instance, when news agencies use GenAI to automate copyediting and article writing, they might reassign journalists to serve as information curators or trainers for GenAI programs. Reassignments like this could then put pressure on journalism programs to adjust their curricula to ensure that their aspiring journalists are prepared for their profession (Frey & Osborne, 2023). When employers start requiring GenAI skills and experience, and university degree programs begin teaching discipline-specific GenAI skills, students develop a strong incentive to become proficient with GenAI tools. If university students do not practice using GenAI programs in their academic studies, they will likely be at a competitive disadvantage when seeking professional employment.

Trust is one key predictor of whether students will use technology, such as GenAI programs. Students tend to trust and therefore use technology more when it is not overly complicated and when it serves an important function in their classwork. However, with GenAI programs specifically, the more significant challenge in developing trust is arguably the ability to evaluate the veracity of GenAI outputs, e.g., by identifying authorship. When reading traditional books and articles, students can see the authors' credentials. These credentials can signal a scholarly work's credibility and accuracy. GenAI programs are not human and therefore do not possess credentials.

The purpose of this manuscript is to provide a framework capable of yielding working hypotheses focusing on how university students can build trust in GenAI programs and their outputs. This framework involves adopting the authors' critical trust framework and applying it within critical friend groups. This manuscript begins with a background on GenAI technology and its applications, followed by an examination of existing frameworks for explaining the development of trust in GenAI. It then lays the foundation for the concept of critical trust, its application in critical friend groups, and its use for building trust in GenAI programs and outputs. The manuscript concludes with a series of applications of critical trust in critical friend groups.

Background on GenAI

Generative AI (GenAI) systems, including large language models (LLMs), generate text by learning statistical regularities from large datasets and producing the most likely next words in context. While built on advanced machine-learning architectures, the technical details are less important here than the practical implication: GenAI produces probabilistic outputs that can appear coherent and reasoned without possessing human understanding or intention. This

distinction is central to determining when and under what conditions trust in AI-generated content is warranted.

The technology powering ChatGPT and others, such as Microsoft's Copilot and Google's Gemini, was built on artificial intelligence (AI) innovations from previous decades, originally aimed at improving the efficiency of work tasks in areas such as teaching and learning (Chen et al., 2020). Specific definitions of AI vary, but broadly, AI programs can perform human-like intellectual activities such as categorization, inference, error analysis, and problem-solving (Chen et al., 2020). Increasingly, AI systems can leverage their human-like qualities to collaborate with users on essential tasks, which in recent decades have included responding to learners' queries and adapting their services and support to meet learners' individual needs (Ouyang & Jiao, 2021). In fact, the collaboration between the human user and GenAI technology, in some respects, mirrors human interactions across various contexts (Ouyang & Jiao, 2021).

Current GenAI Programs and Capabilities

The list of existing GenAI programs is extensive, including Google Gemini, Microsoft Copilot, Synthesia, and Dalle-E, and continues to grow by the month. GenAI programs currently enable users to create unique text, images, videos, code, sound, and other content, such as diagrams and 3D renderings (Banh & Strobel, 2023). For instance, by entering a simple text prompt, student users of the free GenAI program Google Gemini can create simple stock photos for class presentations or imaginative images (e.g., futuristic cities or mythical animals) to fuel engagement in literature. GenAI program users can also translate media in the following ways: Text to Text; Text to Image; Text to Audio; Image to Text; Image to Image; Text to Video; Text+Video to Video; Video to Video; Image+Text to Image; Text to 3D Image; and Text to 3D

Animation (Bandi et al., 2023). These functionalities enable students, for example, to create a textual description of complex images, such as charts and graphs, or even a video for a historical documentary project by entering a simple textual prompt.

Other GenAI programs can produce high-quality concept art, music, and animation (Epstein et al., 2023; Fui-Hoon Nah et al., 2023; Yu & Guo, 2023). Programs such as Suno and Aiva enable student users to create and modify original music. For instance, students taking a music appreciation course could use these GenAI programs to convert a symphony from the classical period into an earlier baroque piece, bringing the differences between the two styles into sharper relief. In addition, the GenAI program Midjourney can help guide student users in creating their own unique visual and graphic art (Chiu, 2023; Epstein et al., 2023; Fui-Hoon Nah et al., 2023; Kanders et al., 2024). For example, business students creating an advertising campaign could upload a basic image of their product to one of these GenAI programs to add a background, setting, and characters, thereby making their product more appealing to potential customers.

In addition, GenAI programs can provide students with academic supports such as personal tutoring, individual feedback to correct errors, dialectic exercises, and practice responding to ethical quandaries with a chatbot (Ahmad et al., 2023; Baidoo-Anu & Ansah, 2023; Chiu, 2023; Cooper, 2023; Fui-Hoon Nah et al., 2023; Hu, 2024; Hwang & Chen, 2023; Rashel et al., 2024; Su & Yang, 2023; van den Berg, 2024; Yu & Guo, 2023). For instance, a humanities major taking their required statistics or Calculus course could use a GenAI program like ChatGPT to obtain feedback on their completed math homework, thereby improving their performance on independent problem sets. Such applications could help students independently improve their understanding and application of course concepts in their future careers. This

adaptive assistance is possible when GenAI programs collect and analyze data while they interact with the student (Chiu, 2023; Fui-Hoon Nah et al., 2023; Kaplan-Rakowski et al., 2023; Rashel et al., 2024).

Finally, GenAI programs can facilitate students' brainstorming and creative writing process for deliverables such as outlines, summary notes from research articles, and even mind maps that illustrate connections between concepts and ideas (Ali et al., 2024; Ahmad et al., 2023; Chiu, 2023; Cooper, 2023; Fui-Hoon Nah et al., 2023; Kanders et al., 2024; van den Berg, 2024; Yu & Guo, 2023). Using GenAI programs in this way could help students improve productivity and avoid missed deadlines due to writer's block. Such support could be especially impactful for special populations such as students with disabilities and English language learners. For instance, students with learning disabilities could use GenAI programs to summarize complex texts in different writing styles with more accessible vocabulary (Cooper, 2023; Pack & Maloney, 2023). GenAI programs could even provide translations of class resources, such as videos and writing scaffolding, to support different levels of language proficiency (Cooper, 2023; Pack & Maloney, 2023; van den Berg, 2024).

Reliability of GenAI Outputs

Although GenAI technology promises improvements in student learning, its limitations are notable. By far the most concerning limitation is the presence of hallucinations, which undermine the reliability of GenAI outputs. Hallucinations in GenAI programs refer to false "facts" that distort a person's sense of reality (Maleki et al., 2024). GenAI hallucinations stem from two primary, generalizable issues: the complexity of models and algorithms for processing data and the use of incomplete or biased training data (Gondode et al., 2024). GenAI hallucinations also result from bias in model design, mismatches in heuristic data collection,

incorrect encoding of input data, and incorrect decoding of output data (Siontis et al., 2024). Studies such as those by Kim et al. (2025) in Bioengineering have developed tools to address the causes of hallucinations in GenAI. However, these tools are still in early development and, as such, are imperfect. Meanwhile, other studies explore the utility of AI literacy for identifying GenAI hallucinations and mitigating their adverse effects, placing greater responsibility on users to detect them than on improving the technology (Hien & Nga, 2024; Walter, 2024). However, the impact of AI literacy on the detection and rejection of GenAI hallucinations depends on users' critical focus on error detection.

Models for Trust in GenAI

Trust generally develops from the predictability of outcomes based on patterns in past experiences. For instance, past GenAI hallucinations could undermine a student's confidence in GenAI's reliability, thereby eroding their trust in GenAI outputs. Therefore, it is crucial to understand how student users develop the trust necessary to engage with and therefore benefit from GenAI outputs. Understanding how their trust develops could also indicate how to rebuild it after exposure to GenAI hallucinations.

The Technology Acceptance Model

The foremost framework for understanding trust in technology, such as GenAI programs, is the Technology Acceptance Model (TAM). TAM explains the factors affecting users' willingness to engage with and adopt new technology. TAM has various versions but is based on two key factors: perceived usefulness and perceived ease of use, which have been shown to significantly influence the acceptance of new technology (Granic & Maragunic, 2019). Over time, TAM has been integrated with other theories to create the Unified Theory of Acceptance and Use of Technology, which incorporates elements of the Theory of Reasoned Action and

Social Cognitive Theory. These combined theories emphasize the importance of usefulness and reliability, indicating that when individuals assess their trust in new technology, they primarily consider its practical aspects.

In addition, Glikson and Woolley (2020) conducted a comprehensive literature review of trust in AI (prior to the release of GenAI in 2022) across multiple disciplines and identified a connection between trust in AI, the type of AI, and its capabilities. They found that trust in AI is influenced by several technological characteristics, including tangibility, transparency, reliability, immediacy, and, most notably, the AI's anthropomorphic qualities (Glikson & Woolley, 2020). Their findings suggest that people are more likely to trust AI when it functions as expected and exhibits human-like traits in its appearance, voice, and behavior.

Interpersonal Trust

The findings in Glikson and Woolley (2020) highlight the multifaceted nature of trust in AI programs, extending beyond functional aspects to include their anthropomorphic qualities. This suggests that human-like attributes—such as appearance, the tone of voice, geographic-related accents, and communication style—have a significant impact on the level of trust individuals place in GenAI programs (Glikson & Woolley, 2020). Therefore, it is helpful to reconceptualize trust in GenAI as a hybrid construct that merges elements of technological trust with dimensions traditionally associated with interpersonal trust. Interpersonal trust refers to trust developed through interactions among two or more human beings. By analyzing trust through both technology acceptance frameworks, such as the TAM, and psychological theories of human trust development, researchers can gain a more comprehensive understanding of the factors that shape trust in GenAI programs and outputs.

The psychological theories of human trust development, most notably, include the concepts of selective and social trust. Harris et al. (2012) argue that children start to engage in *selective trust* in early childhood, between the ages of three and seven. Selective trust refers to the tendency for children to disbelieve information from unfamiliar sources if it contradicts what they have heard from familiar individuals, such as their parents, siblings, and teachers (Harris et al., 2012). Children in this early childhood age group also form *social trust* when the people in their lives keep promises (Markson & Luo, 2020). When children reach adolescence, they continue to develop social trust by assessing knowledgeability based on reputation (Lee et al., 2015). However, at this stage, children also evaluate intentions to determine trustworthiness (Fett et al., 2014). During adolescence, children's trust is also associated with a variety of interdependent factors, such as the risk of negative outcomes from trusting someone, and children's individual differences in empathic abilities, impulsivity, and antisocial tendencies (van de Groep et al, 2020). Due to this increased complexity, adolescents tend to have lower overall levels of trust than young children, a pattern that persists into adulthood (Flanagan & Stout, 2010).

This hybrid construct of technological and interpersonal trust appears to be a new idea and is therefore not represented in the literature reviewed by the authors of this manuscript. The authors found only one study by Leighton et al. (2016), which suggested that students' trust in information is stronger when they learn the source is human. The absence of this trust construct in existing literature may be linked to a lack of anthropomorphic qualities in past AI programs and to the perception that human-like AI is a fantasy created by science fiction writers.

Critical Trust Framework

The development of social trust during adolescence and the tendency for students to trust information more when it comes from a human being create an environment inhospitable to trusting GenAI programs and outputs. The environment becomes even more problematic for students from historically marginalized communities, such as students of color. Many of these students come from communities that have endured systemic discrimination involving unethical uses of technology (e.g., medical 'treatments' during the Tuskegee Study, 1932-1972) and inequitable access to critical communications technology such as the internet and cell towers for smart devices. Therefore, in this section, the authors developed a new framework of critical trust to explore how students from marginalized communities might navigate through their potentially stronger distrust of GenAI programs and outputs.

Background on Systemic Racism

Brown et al. (2025) define systemic racism as the unfair distribution of benefits by race from economic and social policies and institutions, as well as unequal political power, legal rights, cultural influence, and psychological health. Some manifestations of systemic racism include over-policing in majority minority neighborhoods, punishments for drug offenses that are harsher than usual, and racial gerrymandering. This unfair system began prior to the American Revolutionary War, with the arrival of the first enslaved Africans in Jamestown, Virginia, in 1619. Legislators then embedded systemic racism into American law and culture for nearly 250 years until after the U.S. Civil War, and it persisted during and well after Reconstruction. It persisted through Jim Crow laws, despite ratification of the 14th Amendment, the Civil Rights Act (1964), and multiple Supreme Court cases affirming race protections, including *Brown v. Board of Education* (1954).

Critical Race Theory (CRT) emerged among a group of legal scholars to help teach the public about systemic racism. According to Delgado and Stefancic (2023), CRT asserts that race is a social construct, as evidenced by the ever-changing racial categories and their sociopolitical influences. As such, social differences between racial demographics depend on structures within social systems (e.g., the U.S. Census). Furthermore, CRT asserts that racism is pervasive in the United States throughout many aspects of social life, and that White participation in movements to eliminate racism has historically occurred only when it also benefited White people (Delgado & Stefancic, 2023). Finally, CRT promotes counternarratives that incorporate the perspectives and lived experiences of people of color who have endured discrimination throughout U.S. history (Delgado & Stefancic, 2023).

Discrimination in Information Technology

Unequal access to high-speed broadband internet is a significant form of systemic racism that hinders many historically marginalized communities from keeping pace with and developing trust in information technology. Access to broadband, especially high-speed broadband, is critical for operating high-tech programs such as those involving GenAI. Since the introduction of this technology in the early 2000s, higher-speed broadband internet access has developed first and more rapidly in wealthy urban centers like Silicon Valley, New York City, Austin (Texas), and Seattle (Graves et al., 2021). These urban centers are home to some of the world's largest technological corporations. Broadband access has now reached every major urban center in every U.S. state, as well as most suburban and rural areas, but there are persistent differences in broadband access and internet speeds by neighborhood within urban, suburban, and rural areas (Graves et al., 2021). For instance, a 2017 study involving 905 large U.S. cities found that 30%

of urban households lacked access to high-speed broadband internet (Li et al., 2023). The researchers found that broadband access was 15% lower than average in majority-Black and Hispanic neighborhoods (Li et al., 2023). Since then, the overall gap in access to high-speed broadband internet, initially 30%, has decreased to 24%, but the role of race and income in this gap has persisted (Li et al., 2023).

An even more concerning and insidious form of discrimination is the data-sharing practices in high-tech companies, especially with third-party vendors. Companies often sell customer data to third-party vendors (Abrardi et al., 2024). Sharing data with third-party vendors can make users vulnerable to security breaches by hackers aiming to steal user data and passwords to sell on the dark web to scammers (Ke & Sudhi, 2023). Theft of personal data can cause problems for all users. It is particularly challenging for many historically marginalized communities that lack the resources to repair damage to their reputations and finances (Sannon & Forte, 2022).

Additionally, companies share data with government agencies and law enforcement when requested (Lamdan, 2024). For instance, federal and state law enforcement agencies can request data on suspects, including their search history and GPS locations (Lamdan, 2024). This data sharing could be especially problematic for people from historically marginalized communities. Government and law enforcement's access to user data could make it easier, for instance, to target and harass people of color for crimes they did not commit (Lamdan, 2024). People of color have a history in the U.S. of unequal treatment by law enforcement and of wrongful convictions (Lamdan, 2024).

From Critical Hope to Critical Trust

Since the release of Paulo Freire's Pedagogy of the Oppressed in 1968, many researchers have explored ideas of self- and collective liberation from oppressive systems, such as those involving technology. Freire argued that liberation involves action and reflection on the world to transform it (Freire, 2020). Freire's ideas are now encapsulated (for educators) in the concept of *critical hope*. The most recent and extensive exploration of critical hope in education occurred over the past decade. The first, *Discerning Critical Hope in Educational Practices*, was published in 2014. Several chapters in this book address the concept of critical hope specifically among people of African ancestry throughout the Western world, following centuries of colonialism. The second is a book entitled *Critical hope: How to grapple with complexity, lead with purpose, and cultivate transformative social change*. Chief among the author's seven principles of critical hope is the need to practice truth-telling, ask critical questions publicly of those in power, interrupt dynamics of oppressive systems, and invite marginalized communities to the decision-making table (Grain, 2022).

Critical trust does not offer comforting narratives that gloss over suffering or hardship, nor does it rely on the myth that individual effort alone can overcome systemic barriers. Furthermore, critical hope does not promote the naive belief of classical liberalism that historical injustices can be eradicated with the stroke of a pen. This naïve belief has led to great disappointment among many people in the U.S. who hoped for an end to discrimination through the passage of the Civil Rights Act in 1964 and the Voting Rights Act of 1965. Critical hope, in contrast, encourages individual action to bring about social change. The *hope* part of critical hope is the belief that an individual can make incremental changes to eliminate injustices. Therefore, critical hope rejects fatalism that encourages marginalized communities to abandon the social institutions that they paid for through their taxes.

Like critical hope, *critical trust* requires recognizing past false promises of social progress made by lawmakers and other powerful figures. Recognition of past false promises creates necessary doubt in the altruistic motivations of those in power who claim to stand for social progress. Doubt of motivations then creates the imperative to work outside the systems of power through individual action. Of course, like critical hope, critical trust requires the belief that it is possible to work outside of the systems of power to create social change, thereby avoiding the despair about which Freire warned his readers. Working outside of existing power systems takes two forms. The first must be a person's efforts to subvert within their own sphere of influence the mechanisms by which those in power perpetuate historical injustices. The second must be an individual's efforts to organize others into subversive groups that delegitimize the mechanisms that perpetuate historical injustices.

Chopra and Wallace (2003) hold that trust can arise only when there is a state of dependence between the trustor and the trustee, and when acting on this dependence involves some perceived level of risk. Therefore, trust is not just about functionality or accuracy. It is relational, shaped by vulnerability and the possibility of harm. This framing clarifies that trusting GenAI systems is never neutral or risk-free; it demands that we acknowledge our dependence while also recognizing the uncertainty and potential for harm. In this context, critical trust is not about categorically rejecting GenAI or blindly embracing it. Instead, it is about adopting a conscious, reflective stance toward our engagement with GenAI and a distributed ethic of trust to help navigate the emotional, social, and epistemic risks of using GenAI. By embedding critical trust in a web of human accountability, we shift trust in GenAI from an individual act to a collaborative ethical practice—one grounded in relationships, reflection, and resistance to the illusion of machine neutrality.

Critical Trust in GenAI

Critical trust in GenAI programs acknowledges a long and sordid history of misinformation that aimed to deceive historically marginalized communities. For example, critical trust is mindful of the U.S. government's lies to Black men from Alabama who participated in the Tuskegee Experiment (Jones, 2012). The government told these men they were being treated for bad blood when, in fact, they were being monitored for the natural progression of untreated syphilis (Jones, 2012). Being mindful of historical misinformation should put people from historically marginalized communities on alert for disinformation generated by GenAI programs. Unlike hallucinations, disinformation involves the intent to deceive using false information. Disinformation appears in deep-fake text, images, and videos. What makes disinformation in GenAI most pernicious is the variety of malign agendas that could be driving it, as well as the systematic messages it could convey to students.

However, the trust component of critical trust acknowledges that a GenAI output could be at least partially accurate and therefore worthy of evaluation, enabling GenAI users to continue engaging with and benefiting from the technology. More importantly, the potential for misinformation in GenAI outputs creates a social imperative for users to implement their own fact-checking processes and advocate for improvements to the technology that filter out false information.

Suspending Disbelief – A Mechanism for Engaging in Critical Trust

Continuing to engage with GenAI programs despite hallucinations and disinformation contradicts users' natural responses to false information. Recall the TAM framework discussed earlier, which states that users would likely reject GenAI programs that did not produce outputs with reliable and, as a result, useful information. Therefore, continuing to engage with GenAI

programs, aware of potential hallucinations and disinformation, requires users to temporarily suspend their disbelief and conduct a good-faith evaluation of GenAI outputs.

Coleridge (1907) first coined the phrase *"*suspension of disbelief*,"* *also known as "*poetic faith*,"* to describe how readers can lose themselves in a fictional story (Kauvar, 1969). However, current work on suspending disbelief is rooted in Gerrig's (1999) distinction between unsystematic and systematic belief, which Holland (2008) characterizes as "a matter of degree." The term unsystematic refers to what a person perceives through their five senses, whereas systematic involves forming a belief after a thorough, reasoned assessment of reality (Holland, 2008). Gerrig's (1999) concept of unsystematic belief is supported by recent findings in neuroscience, which indicate that perceptions form in multiple brain centers in response to environmental stimuli (Holland, 2008). Recent neuroscience literature also aligns with Gerrig's (1999) concept of systematic belief, highlighting the prefrontal cortex's role in using logic and reason to evaluate perceptions formed from sensory information (Holland, 2008).

What these frameworks suggest is that suspending disbelief in GenAI involves a person temporarily remaining in a state between unsystematic and systematic belief and thereby avoiding a rush to conclusions. Suspending disbelief is therefore not a cognitive state of belief in unevaluated information, but instead a choice to postpone an expected judgment and engage in a process of prolonged consideration (Meylan, 2024). Wagner (2024) characterizes the prolonged consideration involved in suspending disbelief (in part) as a precautionary measure. Moreover, postponing judgment requires active effort because it interrupts a human cognitive tendency to use facts and other information to confirm (or not) preexisting beliefs and biases (Meylan, 2024).

The concept of suspending disbelief has not, to the authors' knowledge, been applied to the evaluation of outputs from GenAI programs. Therefore, this manuscript aims to help fill that

gap by engaging with the literature on the suspension of disbelief as applied in another relatively new technology, digital gaming. Power and Corchnoy (2024) define the suspension of disbelief as a temporary, passive neurological state in which a person can temporarily disregard reason and logic. Ignoring reason and logic is possible because the gamer is in an interactive setting with dynamics such as dialogue that make the unreality feel real (Power & Corchnoy, 2024). These dynamics then foster emotional and psychological availability to the gaming content (Power & Corchnoy, 2024). However, immersion in an interactive digital gaming environment does not lead to a feeling of disembodiment, in which the gamer feels disconnected from their physical body (Geniusas, 2022). On the contrary, Geniusas (2022) and others argue that gamers engage in an embodied dual consciousness, in which they allow themselves to believe the game is real while remaining consciously aware of their self-deception.

Here, Geniusas (2022) bases their argument on the work of phenomenologist Theodor Conrad, who demonstrated that people could shift from an egoic perspective of *here and now* to a non-egoic standpoint of *there and then* (Geniusas, 2022). Conrad argued that shifting between *here and there*, and *now and then,* can occur in the flow of an experience, and that shifting to a *there-and-then* perspective can create displaced experiences, i.e., experiences not directly experienced by the person shifting their perspective (Geniusas, 2022).

Recall that, according to the TAM framework, GenAI users tend to reject technology that can be unreliable. Thus, gamers need to situate themselves between unsystematic and systematic beliefs. The suspension of disbelief that occurs in a digital gaming environment provides a helpful framework for understanding how a GenAI user can engage earnestly with a GenAI chatbot, even when its outputs might be hallucinations or disinformation. Like digital gaming environments, GenAI chatbots give users a false sense of reality through their humanlike

interactions. As with digital gaming environments, interactions between GenAI chatbots and human users could make the unreal feel real. They also encourage users to share their inner thoughts and seek feedback or guidance from GenAI programs, much as they would from another human. However, as in digital gaming environments, users should enter an embodied dual consciousness in which they speak to the chatbot as if it were a human, while remaining aware that it is not. This embodied dual consciousness is critical for users to strike a balance between prolonged and open-minded consideration of GenAI outputs and the need for critical analysis.

Critical Friend Groups

One potential way to balance prolonged consideration of GenAI outputs with the need for critical analysis is to form critical friend groups. A critical friend is a supportive person who uses their objective point of view to challenge the ideas of those close to them (Storey & Wang, 2017). Critical friends challenge each other's ideas by asking and answering uncomfortable questions (Storey & Wang, 2017). The purpose of a critical friendship is not to win an argument but to help a person develop professionally through complete honesty, trust, respect, and shared values (Hardiman & Dewing, 2014). Critical friends gather in groups to share diverse perspectives, thereby further enhancing their professional development (Storey & Richard, 2015).

Background on Critical Friend Groups

The contemporary origins of the concept of critical friends date back to the 1970s in the United Kingdom, when an educational researcher, Desmond Nuttall, specialized in the study of self-appraisal (Heller, 1988). However, researchers at the Annenberg Institute of School Reform at Brown University were the first to apply the critical friends concept to their Critical Friends

Group program (Dunne & Honts, 1998). Higher education researchers have recently begun exploring the Annenberg Institute's critical friends group model and its potential use in postsecondary education. For instance, Storey and Wang (2016) explored the use of critical friend groups in graduate education. These researchers adapted an evidence-based protocol developed by the Annenberg Institute to facilitate critical friend groups in graduate-level educational leadership courses. Similarly, Jones and McNulty (2023) explored critical friend groups within the context of university study abroad programs. Study findings revealed that after participating in a study abroad critical friend group, students were more likely to explore different teaching methods and feel competent in teaching diverse students (Jones & McNulty, 2023).

GenAI Critical Trust within Critical Friend Groups

Critical friend groups that engage with critical trust could foster among university students and faculty members a collective skepticism of GenAI outputs. Foundational to these groups would be an acknowledgment of past harms caused by technology. This skepticism creates the imperative to identify and filter out false content when using GenAI outputs for teaching, learning, and research. Collective skepticism could also provide time and space for critical dialogue about acceptable uses to preserve the integrity of teaching, learning, and scholarship.

These critical friend groups would ideally be centered within existing university communities. University communities include classrooms, student study groups, student organizations, informal social networks, etc. Such centering reflects an orientation to systems change theory, which focuses on eliminating social injustices (Foster-Fishman & Watson, 2017). Storey and Wang (2016) and Jones and McNulty (2023) suggest that critical friend groups

require an initiator to engage in critical trust of GenAI outputs. An initiator would be someone who, for example, suspects that a GenAI output is a hallucination or disinformation and proposes the problem to the critical friend group. The initiator would present the group with evidence that justifies their suspicion. The group would then engage in a dialogue to systematically think through the evidence.

In addition, critical friend groups could engage in dialogue aligned with their objectives, using a specific protocol. A protocol is a series of steps with time limits that provide a logical structure, promote continuous focus on objectives, and maximize participation (McDonald et al., 2015). Protocols enable educators to share feedback in a safe space and analyze problems carefully, considering all their complexities without rushing to judgment (McDonald et al., 2015). One helpful approach to engagement in the protocol could be the use of dialectics. Dialectics emphasizes deepening understanding and fostering positive change in participants and the world around them (Freire, 1970, 1995). Kazepides (2012) argues that dialectics also require outlining degrees of agreement on a specific issue, as well as respect, trust, open-mindedness, a willingness to listen, and a risk-taking approach to one's own preconceptions, fixed beliefs, biases, and prejudices in pursuit of truth. Notably, the characteristics that Kazepides (2012) outlines here are Aristotelian virtues that classical philosophers believed were required to facilitate the use of formal reason and logic.

Finally, critical friend groups rely on six accountability tenets to build trust between members. These tenets include the following: Relational Accountability, Shared Vulnerability, Transparency and Disclosure, Reflection and Dialogue, Ethical Literacy, and Agency and Autonomy. First, members of the critical friend group have relational accountability with one another, involving shared responsibility to engage in good-faith, collaborative analysis of GenAI

outputs. This tenet makes space for honest questioning, collaborative sensemaking, and dialogic feedback about how GenAI is being used, interpreted, or even misused. When trust in GenAI becomes habitual or overly reliant, friend groups can intervene, prompting the user to pause, reconsider, and recalibrate.

Second, when discussing GenAI use with friends, there is space to acknowledge and share vulnerabilities, recognizing that every engagement with GenAI entails some degree of uncertainty, risk, and potential for harm. These vulnerabilities may stem from a lack of understanding of how a system works, overconfidence in its outputs, or unintentionally reproducing biases embedded in the data and design. When entering into accountability conversations with peers, there is an opportunity to share insights. This act of openness is itself a form of resistance to the illusion of mastery that often accompanies technological use. It positions vulnerability not as weakness, but as a site of critical strength, thus a starting point for mutual reflection, ethical awareness, and recalibrated trust.

Third, Transparency and Disclosure requires critical friend groups to share their experiences and vulnerabilities in a safe, protected space. This could involve discussing the capabilities and limitations of GenAI tools, identifying any potential unintended consequences, or sharing moments when the GenAI system failed to meet expectations. Through transparency within the critical trust friend group, the accountability network can offer guidance on how to adjust expectations or behavior in future interactions, thus pushing into the fourth tenet, Reflection and Dialogue. Building trust in GenAI is an iterative process. Students commit to ongoing dialogue with their peers about the evolving role of GenAI in their lives. The fifth tenet of GenAI critical trust is the development of ethical literacy. This involves recognizing, questioning, and responding to the moral dimensions of GenAI systems, including concerns

about data privacy, algorithmic bias, surveillance, labor, and the broader socio-political impacts of automation.

The last tenet relies on scholarly autonomy and agency. At the heart of critical trust in critical friend groups is the imperative that trusting GenAI should not equate to surrendering personal agency. Unlike passive reliance, critical trust demands active, reflective control over decision-making processes. It positions GenAI not as an authority, but as a tool. Critical friend groups play a crucial role in sustaining this stance. Through regular dialogue and mutual reflection, friends can help reinforce the idea that GenAI should augment, not replace, users' reasoning.

Implications

Critical friend groups are circles of students who engage in critical trust—offering honest feedback, asking hard questions, reflecting deeply, and supporting each other's academic growth. Engaging in critical trust within critical friend groups could potentially help prevent university students from using misinformation and disinformation from GenAI hallucinations in their classwork. Preventing the use of misinformation and disinformation is especially critical for students from historically marginalized communities, who would suffer the most damage to their academic reputations. Critical trust in critical friend groups could also help students evaluate the appropriateness of their GenAI use.

In the classroom

The nexus of critical friend groups should be in the classroom, with instructors modeling critical trust for their students. Therefore, course syllabi should clearly communicate the purpose and expectations of critical friend groups. This includes outlining norms for respectful dialogue, the structure of peer feedback, and the value placed on collaboration. Relationship building

should be an integral part of the course's foundation. For example, the first few weeks of classes could include community-building activities, shared reflections on identity and values, and the collaborative creation of a class social contract. Critical trust groups can operationalize this relational anchoring by having students collaboratively analyze and evaluate GenAI outputs and discuss the risks and benefits of GenAI use. Faculty members can facilitate this by creating structured opportunities for students to reflect together on how and why they turn to GenAI tools and to critically evaluate the influence of those tools on their thoughts, writing, and decision-making.

In addition, assignments should be thoughtfully designed, balancing group work and individual accountability. For example, a significant project might include checkpoints where critical friend groups offer structured feedback using a protocol, such as "I Like, I Wonder, I Suggest," followed by individual reflection papers. Class discussions, too, should model dialectic learning, where students feel empowered to push back, ask questions, and offer critical feedback. In the rapidly evolving landscape of higher education, the framework of critical trust challenges faculty and students to resist passive acceptance of GenAI outputs and instead cultivate reflective, relational, and ethically grounded engagement with these tools.

Faculty can model critical trust by being transparent about when and how they use GenAI in their teaching (e.g., syllabus design, assessment feedback) and by openly discussing the limits of those tools. Instructors can also design assignments that require students to make their decision-making processes visible, explicitly distinguishing between their own reasoning and the contributions of GenAI. Peer groups can then provide feedback not only on academic content but on the integrity and intentionality of GenAI use, reinforcing a culture of reflection.

In Academic Scholarship

Student scholarship also demands transparency, and peer feedback is essential when utilizing GenAI programs. Scholarship includes published work in various venues, including colloquia, symposia, academic competitions, internal and external journals, and master's and doctoral theses and dissertations. The thought, research, and writing involved in scholarship take extensive time and effort. As demonstrated in this manuscript, many GenAI applications support idea generation, information access, and writing improvement. Therefore, GenAI could create the temptation for students to use it to reduce the time and effort required to think, research, and write when generating publications. When this temptation to use GenAI to save time and effort is coupled with indiscriminate trust, students open themselves up to accusations of fraud, either by negligently spreading misinformation or by plagiarizing. The transparency and peer feedback promoted in critical friend groups can help students prevent accusations of academic fraud.

However, for critical friend groups that practice critical trust to gain a foothold in student research at the university level, the research structure might need to change. Take, for example, the standard undergraduate capstone project, master's thesis, and doctoral dissertation. These are projects that demonstrate an individual student’s cumulative knowledge and skill related to their discipline and specific research topic. Projects begin with the individual student stating a problem and a research question. The student then conducts a literature review and, if applicable, an empirical study, followed by an independent discussion of the findings. Students could use GenAI applications to support any stage of this process. However, without collaboration during this individual research process, students lack external checks on the ethics of GenAI use and the veracity of its outputs. Critical friend groups can provide opportunities for these collaborative external checks.

University faculty members can provide students with a structure for scholarly research that involves critical friend groups. For instance, faculty members could model critical friend conversations when preparing their students for collaborative group work in class projects. This model could draw on the principles outlined in this manuscript to guide the development of lists of questions to facilitate critical friend discussions within these project groups. In addition, faculty members could require engagement in critical friend groups and a GenAI-use section in their publications. They could require students to write narratives about how they used GenAI in their research, the interactions they had in critical friend groups to ensure the ethical use of the technology, and the accuracy of GenAI outputs.

In Professional Growth Opportunities

Transparency and peer feedback in critical friend groups are equally important when using GenAI programs to pursue student fellowships and internships. Fellowship applications typically require a cover letter, a statement of purpose, a resume, and letters of recommendation. Fellowships also often involve a capstone deliverable requiring research. Internships have a similar application process and a series of duties frequently requiring research, analysis, problem-solving, and writing. Historically, these professional opportunities were critical for fellows and interns to demonstrate their knowledge and skills to future employers. However, widespread access to GenAI programs enables fellows and interns to use chatbots to perform this work on their behalf, producing deliverables such as research projects and papers. As a result, mistakes in a student's deliverables for a fellowship or internship could tarnish their reputation and potentially harm their prospects for future employment.

Critical friend groups could provide fellows and interns with an essential source of transparency and peer feedback. Career service departments on university campuses, which

assist students seeking fellowships and internships, could be ideal places to create critical friend groups to shepherd students through the application process. However, students could initiate critical friend groups at their fellowship and internship placements. These onsite critical friend groups could include other interns, fellowship and internship mentors, and supervisors. Mentors and supervisors would be more likely to have specific knowledge to effectively critique the information in fellows' and interns' deliverables.

In Student Life

Finally, within student affairs and residence life, GenAI has the potential to personalize engagement through targeted newsletters and virtual assistants that can respond to common inquiries about quiet hours, maintenance requests, or community expectations. More advanced tools may analyze language in anonymous feedback or usage data to flag potential concerns related to student well-being, enabling earlier interventions or support referrals. To facilitate critical trust in these settings, resident assistants and hall directors could build critical friend groups into their existing programming. Critical friend groups could play a crucial role in fostering critical trust in these settings. These groups, comprising student leaders, residence life staff, and advisors, could periodically review GenAI outputs for accuracy, tone, cultural responsiveness, and potential bias. They could also provide a structured space for ethical reflection on the use of GenAI programs, making engagement in critical trust within critical friend groups part of a general approach to campus life, rather than a mere academic exercise.

References

Abrardi, L., Cambini, C., & Pino, F. (2024). Regulating data sales: The role of data selling mechanisms. *Telecommunications Policy, 48*(8), 102813.

Ahmad, N., Murugesan, S., & Kshetri, N. (2023). Generative artificial intelligence and the education sector. *Computer, 56*(6), 72–76.

Ali, S., Ravi, P., Williams, R., DiPaola, D., & Breazeal, C. (2024). Constructing dreams using generative AI. In *Proceedings of the AAAI Conference on Artificial Intelligence, 38*(21), 23268–23275.

Alomary, A., & Woollard, J. (2015, November 21). How is technology accepted by users? A review of technology acceptance models and theories. In *Proceedings of the IRES 17th International Conference* (pp. 1–4). https://eprints.soton.ac.uk/382037/1/110-14486008271-4.pdf (ePrints Soton)

American Civil Liberties Union. (2025, May 1). *ICE deports 3 U.S. citizen children held incommunicado prior to the deportation* [Press release]. https://www.aclu.org/press-releases/ice-deports-3-u-s-citizen-children-held-incommunicado-prior-to-the-deportation

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52–62.

Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The power of generative AI: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet, 15*(8), 260.

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets, 33*(1), 63.

Bordum, A. (2004). *Trust as a critical concept* (Working Paper No. wp2004-004). Copenhagen Business School. https://research-api.cbs.dk/ws/portalfiles/portal/59000504/wp2004_004.pdf

Bozalek, V., Leibowitz, B., Carolissen, R., & Boler, M. (Eds.). (2014). *Discerning critical hope in educational practices*. Routledge.

Brown, T. H., Lee, H. E., Hicken, M. T., Bonilla-Silva, E., & Homan, P. (2025). Conceptualizing and measuring systemic racism. *Annual Review of Public Health, 46*(1), 69–90.

Canaan, J. (2005). *Developing a pedagogy of critical hope*. *Learning & Teaching in the Social Sciences, 2*(3), 159–174.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278.

Chiu, T. K. (2023). The impact of generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*. Advance online publication, 1–17.

Chopra, K., & Wallace, W. A. (2003). Trust in electronic environments. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS-36)*. https://doi.org/10.1109/HICSS.2003.1174902

Černý, M. (2022). The history of chatbots: The journey from psychological experiment to educational object. *Journal of Applied Technical and Educational Sciences, 12*(3), 322–322. https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai

Chegg, Inc. (2025, January 28). *Chegg Global Student Survey 2025: 80% of undergraduates worldwide have used GenAI to support their studies—but accuracy a top concern* [Press release]. https://investor.chegg.com/Press-Releases/press-release-details/2025/Chegg-Global-Student-Survey-2025-80-of-Undergraduates-Worldwide-Have-Used-GenAI-to-Support-their-Studies--But-Accuracy-a-Top-Concern/default.aspx

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*(3), 444–452.

Costa, A. L., & Kallick, B. (1993). Through the lens of a critical friend. *Educational Leadership, 51*, 49–49.

Delgado, R., & Stefancic, J. (2023). *Critical race theory: An introduction* (Vol. 87). NYU Press.

Duncan-Andrade, J. (2009). Note to educators: Hope required when growing roses in concrete. *Harvard Educational Review, 79*(2), 181–194. https://doi.org/10.17763/haer.79.2.nu3436017730384

Duncan-Andrade, J. M. R. (2010). *What a coach can teach a teacher: Lessons urban schools can learn from a successful sports program*. Peter Lang.

Dunne, F., & Honts, F. (1998, April 13–17). That group really makes me think! Critical friends groups and the development of reflective practitioners [Conference presentation]. *Annual Meeting of the American Educational Research Association*, San Diego, CA, United States.

Elena-Bucea, A., Cruz-Jesus, F., Oliveira, T., & Coelho, P. S. (2021). Assessing the role of age, education, gender and income on the digital divide: Evidence for the European Union. *Information Systems Frontiers, 23*, 1007–1021.

Epstein, Z., Hertzmann, A., Investigators of Human Creativity, Akten, M., Farid, H., Fjeld, J., … Smith, A. (2023). Art and the science of generative AI. *Science, 380*(6650), 1110–1111.

Fett, A. K. J., Shergill, S. S., Gromann, P. M., Dumontheil, I., Blakemore, S.-J., Yakub, F., & Krabbendam, L. (2014). Trust and social reciprocity in adolescence—A matter of perspective-taking. *Journal of Adolescence, 37*(2), 175–184.

Flanagan, C. A., & Stout, M. (2010). Developmental patterns of social trust between early and late adolescence: Age and school climate effects. *Journal of Research on Adolescence, 20*(3), 748–773.

Foster-Fishman, P., & Watson, E. (2017). Understanding and promoting systems change. In M. A. Bond, I. Serrano-García, C. B. Keys, & M. Shinn (Eds.), *APA handbook of community psychology: Methods for community research and action for diverse groups and issues* (pp. 255–274). American Psychological Association.

Freire, P. (2020). Pedagogy of the oppressed. In *Toward a sociology of education* (pp. 374–386). Routledge.

Frey, C. B., & Osborne, M. (2023). Generative AI and the future of work: A reappraisal. *Brown Journal of World Affairs*, 30(1), 1–17.

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research, 25*(3), 277–304.

Geniusas, S. (2022). What is immersion? Towards a phenomenology of virtual reality. *Journal of Phenomenological Psychology, 53*(1), 1–24.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660.

Gondode, P., Duggal, S., & Mahor, V. (2024). Artificial intelligence hallucinations in anaesthesia: Causes, consequences and countermeasures. *Indian Journal of Anaesthesia, 68*(7), 658–661.

Grain, K. (2022). *Critical hope: How to grapple with complexity, lead with purpose, and cultivate transformative social change*. North Atlantic Books.

Grain, K. M., & Land, D. E. (2017). The social justice turn: Cultivating critical hope in an age of despair. *Michigan Journal of Community Service Learning, 23*(1), 1–21.

Graves, J. M., Abshire, D. A., Amiri, S., & Mackelprang, J. L. (2021). Disparities in technology and broadband internet access across rurality: Implications for health and education. *Family & Community Health, 44*(4), 257–265.

Greene, M., & Macrine, S. L. (2020). Teaching as possibility: A light in dark times. In *Critical pedagogy in uncertain times: Hope and possibilities* (pp. 81–94). Springer.

Hardiman, M., & Dewing, J. (2014). Critical ally and critical friend: Stepping stones to facilitating practice development. *International Practice Development Journal, 4*(1), 3-1 – 3-19.

Harris, P. L., Corriveau, K. H., Pasquini, E. S., Koenig, M., Fusaro, M., & Clément, F. (2012). Credulity and the development of selective trust in early childhood. In M. J. Beran, J. L. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of Metacognition* (pp. 193–210). Oxford University Press.

Heller, H. (1988). The advisory service and consultancy. In H. Gray (Ed.), *Management consultancy in schools*. Cassell.

Hien, L. T., & Nga, L. P. (2024). Current status and measures to develop critical thinking for pedagogical students in using generative artificial intelligence. *Vinh University Journal of Science: Educational Science and Technology, 53*. https://doi.org/10.56824/vujs.2024.htkhgd180

Holland, N. N. (2008). Spider-Man? Sure! The neuroscience of suspending disbelief. *Interdisciplinary Science Reviews, 33*(4), 312–320.

Hu, Y. H. (2024). Implementing generative AI chatbots as a decision aid for enhanced values clarification exercises in online business ethics education. *Educational Technology & Society, 27*(3), 356–373.

Hwang, G.-J., & Chen, N.-S. (2023). Exploring the potential of generative artificial intelligence in education: Applications, challenges, and future research directions. *Journal of Educational Technology & Society, 26*(2), 1–18.

Jeffrey, R. C., & Burgess, J. P. (2006). *Formal logic: Its scope and limits*. Hackett.

Jones, J. H. (2012). *Tuskegee's truths: Rethinking the Tuskegee syphilis study*. UNC Press Books.

Jones, M. K., & McNulty, C. P. (2023). Critical friend groups: The unexpected relationships among education majors during study abroad. *Journal of Education and Learning, 12*(4), 14–25.

Kalota, F. (2024). A primer on generative artificial intelligence. *Education Sciences, 14*(2), 172. https://doi.org/10.3390/educsci14020172

Kanders, K., Stupple-Harris, L., Smith, L., & Gibson, J. L. (2024). Perspectives on the impact of generative AI on early-childhood development and education. *Infant and Child Development, 33*(4), e2514.

Kaplan-Rakowski, R., Grotewold, K., Hartwick, P., & Papin, K. (2023). Generative AI and teachers' perspectives on its implementation in education. *Journal of Interactive Learning Research, 34*(2), 313–338.

Kauvar, G. B. (1969). Coleridge, Hawkesworth, and the willing suspension of disbelief. *Papers on Language and Literature, 5*(1), 91–94.

Kazepides, T. (2012). Education as dialogue. *Educational Philosophy and Theory, 44*(9), 913–925.

Ke, T. T., & Sudhir, K. (2023). Privacy rights and data security: GDPR and personal data markets. *Management Science, 69*(8), 4389–4412.

Kim, S., Park, C., Jeon, G., Kim, S., & Kim, J. H. (2025). Automated audit and self-correction algorithm for seg-hallucination using MeshCNN-based on-demand generative AI. *Bioengineering, 12*(1), 81.

Lamdan, S. (2022). *Data cartels: The companies that control and monopolize our information*. Stanford University Press.

Lee, N. C., Jolles, J., & Krabbendam, L. (2016). Social information influences trust behaviour in adolescents. *Journal of Adolescence, 46*, 66–75.

Leighton, G. R., Hugo, P. S., Roulin, A., & Amar, A. (2016). Just Google it: Assessing the use of Google Images to describe geographical variation in visible traits of organisms. *Methods in Ecology and Evolution, 7*(9), 1060–1070.

Li, Y., Spoer, B. R., Lampe, T. M., Hsieh, P.-Y., Nelson, I. S., Vierse, A., … Gourevitch, M. N. (2023). Racial/ethnic and income disparities in neighborhood-level broadband access in 905 US cities, 2017–2021. *Public Health, 217*, 205–211.

Luan, H., & Tsai, C.-C. (2021). A review of using machine learning approaches for precision education. *Educational Technology & Society, 24*(1), 250–266.

Macrine, S. L. (Ed.). (2009). *Critical pedagogy in uncertain times: Hope and possibilities* (pp. 119–136). Palgrave Macmillan.

Maleki, N., Padmanabhan, B., & Dutta, K. (2024). AI hallucinations: A misnomer worth clarifying. In *2024 IEEE Conference on Artificial Intelligence (CAI)* (pp. 133–138). IEEE.

Markson, L., & Luo, Y. (2020). Trust in early childhood. In Janette B. Benson (Ed.), *Advances in Child Development and Behavior* (Vol. 58, pp. 137–162). Academic Press/Elsevier.

McDonald, J. P., Mohr, N., Dichter, A., & McDonald, E. C. (2015). *The power of protocols: An educator's guide to better practice*. Teachers College Press.

Megahed, F. M., Chen, Y. J., Ferris, J. A., Knoth, S., & Jones-Farmer, L. A. (2024). How generative AI models such as ChatGPT can be (mis)used in SPC practice, education, and research? An exploratory study. *Quality Engineering, 36*(2), 287–315.

Meylan, A. (2024). What do we do when we suspend judgement? *Philosophical Issues, 34*(1), 253–270.

Mollick, E. (2022, December 14). ChatGPT is a tipping point for AI. *Harvard Business Review*. https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*, 100020.

Pack, A., & Maloney, J. (2023). Potential affordances of generative AI in language education: Demonstrations and an evaluative framework. *Teaching English with Technology, 23*(2), 4–24.

Power, P., & Corchnoy, S. (2024). The world of gaming and Jungian analysis. *Psychological Perspectives, 67*(1–2), 15–35.

Rashel, M. M., Khandakar, S., Hossain, K., Shahid, A., Kawabata, T., Batool, W., … Rafique, T. (2024). AI in education: Unveiling the merits and applications of ChatGPT for effective teaching environments. *Revista de Gestão Social e Ambiental, 18*(10), e09110.

Roveda, L., Magni, M., Cantoni, M., Piga, D., & Bucca, G. (2021). Human–robot collaboration in sensorless assembly task learning enhanced by uncertainties adaptation via Bayesian optimization. *Robotics and Autonomous Systems, 136*, 103711.

Sampat, B., Mogaji, E., & Nguyen, N. P. (2024). The dark side of FinTech in financial services: A qualitative enquiry into FinTech developers' perspective. *International Journal of Bank Marketing, 42*(1), 38–65.

Sannon, S., & Forte, A. (2022). Privacy research with marginalized groups: What we know, what's needed, and what's next. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2), 1–33.

Senge, P. (1990). Peter Senge and the learning organization. *Dimension, 14*. *[Issue/pages needed]*

Siontis, K. C., Attia, Z. I., Asirvatham, S. J., & Friedman, P. A. (2024). ChatGPT hallucinating: Can it get any more humanlike? *European Heart Journal, 45*(5), 321–323. https://doi.org/10.1093/eurheartj/ehad766

Sira, M. (2023). Generative AI takes centre stage: Revolutionizing productivity and reshaping industries. *System Safety: Human-Technical Facility-Environment, 5*(1), 57–65.

Storey, V. A., & Richard, B. (2015). The role of critical friends in supporting institutional change: Developing a collaborative environment. *Journal of Applied Research in Higher Education, 7*(2), 412–428.

Storey, V. A., & Wang, V. C. (2017). Critical friends protocol: Andragogy and learning in a graduate classroom. *Adult Learning, 28*(3), 107–114.

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education, 6*(3), 355–366.

Taylor, S. N. (2014). Student self-assessment and multisource feedback assessment: Exploring benefits, limitations, and remedies. *Journal of Management Education, 38*(3), 359–383.

van den Berg, G. (2024). Generative AI and educators: Partnering in using open digital content for transforming education. *Open Praxis, 16*(2), 130–141.

van de Groep, S., Zanolie, K., Green, K. H., Sweijen, S. W., & Crone, E. A. (2020). A daily diary study on adolescents' mood, empathy, and prosocial behavior during the COVID-19 pandemic. *PLOS ONE, 15*(10), e0240349.

Wagner, V. (2024). *Committing to indecision: A taxonomy of suspension of judgment*. Routledge.

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*(1), 15.

Wu, Y., Jiang, A. Q., Li, W., Rabe, M., Staats, C., Jamnik, M., & Szegedy, C. (2022). Autoformalization with large language models. *Advances in Neural Information Processing Systems, 35*, 32353–32368.

Yu, H., & Guo, Y. (2023). Generative artificial intelligence empowers educational reform: Current status, issues, and prospects. In *Frontiers in Education* (Vol. 8, Article 1183162). Frontiers Media SA.