

THRESHOLDS IN EDUCATION

since 1975

Volume 48
Issue 2
ISSN: 0916-9641

BUILDING AI LITERACY: CRITICAL APPROACHES & PEDAGOGICAL APPLICATIONS SURFING IN A TSUNAMI

**Invited Guest Editors
Marc Watkins
&
Stephen Monroe
University of Mississippi**

Thresholds in Education

www.academyforeducationalstudies.org/Journals/Thresholds/

Thresholds in Education Foundation Board of Directors 2016

Dr. Kerry Burch, Executive Editor

Dr. Patrick Roberts

Dr. Teresa Wasonga

Dr. Kelly Summers

Sarah Militz-Frielink

Patricia Kee

Aim and Scope: *Thresholds in Education*, now an open access online scholarly journal, is dedicated to the philosophical exploration of programmatic, pedagogical, and curricular renovation as well as new education inquires, theories, and viewpoints. *Thresholds* is a space for authors to engage the philosophical method not only to provide constructive critique of extant policy and practice in education but also to explore fresh ideas and viewpoints that will lay the foundation for innovative and critically oriented schools of the future.

Subscription Information: All past, current, and future issues of *Thresholds in Education* are available at no cost @ www.academyforeducationalstudies.org/Journals/Thresholds/.

Copyright © *Thresholds in Education Foundation & Academy for Educational Studies*.

Authors warrant that they are sole owners of the material included in the paper and that it is original and has never been published in any form previously. Authors grant to the *Thresholds in Education Foundation* and the *Academy for Educational Studies* all rights to the paper including rights of copyright. The *Thresholds in Education Foundation* & the *Academy for Educational Studies*, in turn, agree to grant permission to authors to republish their manuscripts subject only to the condition that *Thresholds in Education* be given proper credit as the original source and the *Academy for Educational Studies* as original publisher.

Thresholds in Education is published annually in the late spring of each year. ISSN 01969541 (EISSN pending approval).

Views expressed herein do not necessarily reflect the views of the editor(s), the *Thresholds in Education* Board of Directors, or the Executive Board of the *Academy for Educational Studies*.

Disclaimer:

In consideration of participating in the media production described herein, contributors, their heirs, executor, administrators, legal representatives and assigns release and forever discharge the *Thresholds in Education* Foundations Board, the *Academy for Educational Studies* and their officers, agents, and employees and all other persons connected with *Thresholds in Education* from any and every claim, demand, action, in law or equity that may arise as a result of their participation in this production.

A special thanks to Western Illinois University graduate assistant, Debbie Amoah, who assisted with this issue.

Building AI Literacy: Critical Approaches & Pedagogical Applications Surfing in a Tsunami

Table of Contents

Volume 48, Issue 2

ISSN: 0916-9641

Issue Editors: Marc Watkins & Stephen Monroe

<i>Building AI Literacy: Critical Approaches & Pedagogical Applications Surfing in a Tsunami</i> Marc Watkins & Stephen Monroe	137
<i>Building Critical AI Literacy: An Approach to Generative AI</i> Kathryn Conrad & Sean Kamperman	142
<i>Characterizing ChatGPT's Feedback for FYW: Analyzing Feedback Responses to Inquiry-driven Essays</i> Kirkwood Adams & Maria G. Baker	159
<i>AI & Data Competencies: Scaffolding holistic AI literacy in Higher Education</i> Kathleen Kennedy & Anuj Gupta	182
<i>Collaborative Intelligence: Towards Practical, Critical & Cooperative Teaching & Learning with AI</i> Amy Walter	202
<i>The Stakes Are High. Are the Benefits Bountiful?: HBCU Students, AI, & the Power of Composition</i> Adrienne Carthon	224
<i>Algorithmic Foucault: Digital Feminism, the Panopticon, & the Role of AI in Shaping Musical Identities & Pedagogies</i> Mila Zhu.....	240
Contributors	263

Building AI Literacy: Critical Approaches & Pedagogical Applications Surfing in a Tsunami

Marc Watkins & Stephen Monroe

Abstract

We begin our introduction by acknowledging the valid anxieties of faculty who face rapid technological change brought on by Generative AI (GenAI) tools without adequate institutional support or training. While some scholars advocate for GenAI resistance and others for wholesale adoption, the voices included within this volume argue for a balanced, pragmatic approach that emphasizes transparency, ethical usage, and thoughtful engagement with these powerful new tools. Despite concerns about GenAI's limitations and potential negative impacts, we recognize its remarkable capabilities and argue that higher education is uniquely positioned to critically engage with this technology through its diverse disciplinary perspectives and research methods. In seven collected essays, educators move beyond initial worries to explore creative pedagogical applications that can enhance student learning while maintaining academic integrity.

Keywords: *Generative Artificial Intelligence, Writing, Higher Education, Pedagogical Innovation, Academic Integrity, Faculty Development, AI Literacy*

My advice to you all is to ride the wave...Each and every one of you has a reason to use this technology. If you're an artist, a teacher, a physician, a business person, a technical person, if you're not using this technology, you're not going to be relevant compared to your peer groups and competitors...Adopt it, and adopt it fast.

Eric Schmidt, May 2025

This is a time of intense technological change: an AI wave. An AI tsunami? Eric Schmidt's advice—to teachers and to everyone else—is to lean in, learn, and adopt AI as quickly as possible. He warns that we ignore AI at our peril. It sounds like Silicon Valley hype, but Schmidt is probably right. OpenAI's ChatGPT now has over 400 million weekly worldwide users (Barnum & Seetharaman, 2025). That's impressive growth for a tool that launched little more than two years ago. And, of course, OpenAI is only one company causing wakes. Anthropic, Microsoft, Meta, X, and Schmidt's own Google are also thrashing ahead, delivering new models, recruiting millions of

users, and promulgating the many potential benefits of Generative AI. The resulting waves—in education and in society—are unprecedented and gaining speed.

But tsunamis (and even smaller waves) can be destructive. As educators, we are right to ask skeptical questions and to worry about potential risks. For example, how do these tools threaten student learning? Where is the line between AI assistance and AI cooption? What are we losing in our classrooms, if we adopt, adopt, adopt? Some worrisome hints can be found in studies conducted by the AI companies themselves.

In 2025, Microsoft examined GenAI usage in the workforce by surveying over 300 knowledge workers about how they used Microsoft Copilot. The authors found an alarming connection between heavy usage of GenAI and a negative impact on critical thinking skills. The authors offer a thoughtful warning: “While GenAI can improve worker efficiency, it can inhibit critical engagement with work and can potentially lead to long-term overreliance on the tool and diminished skill for independent problem-solving” (Lee et al., 2025). Their suggestion was not to ban the tool or deny workers access, but to ensure GenAI interfaces are designed to support critical thinking—not to offload it altogether.

Anthropic’s Education Report examined one million anonymized student chats to see how university students use their flagship AI model, Claude. The results are quite surprising. Some of the heaviest users were in STEM fields and nearly half of the top chats were students using GenAI to seek direct answers or shortcuts, instead of using the model as a tutor. Dominant were queries like “solve and explain statistics problems,” “answer earth science questions,” and “solve chemistry problems.” The authors of the study provided a nuanced understanding of student usage: while some students used Claude as a tutor and collaborative partner, many more sought to quickly offload their intellectual labor. The authors warned, “there are legitimate worries that AI systems may provide a crutch for students, stifling the development of foundational skills needed to support higher-order thinking” (*Anthropic Education Report: How University Students Use Claude*, 2025). To that end, Anthropic’s new offering, Claude for Education, introduces “learning mode,” which uses GenAI to guide discovery and Socratic inquiry rather than simply providing students with a speedy answer (*Claude for Education / Partnering with Universities on Responsible AI*, 2025). This is good. We should recognize when a GenAI developer responds to troubling trends in their own research and makes an effort to address those concerns through redesign. Such efforts make our job as educators easier and prove that AI companies and teachers can work together on behalf of our common “customers.” Such tailored AI products show that is possible to encourage students to pursue ethical and transparent GenAI usage, while also promoting learning.

But not all developers are spending such time on human learning problems or creating throttled interfaces, which may or may not prove popular/profitable. Thus far, OpenAI’s response to concerns about human learning has been muted. In lieu of redesigning their interface for students in their ChatGPT Edu offering, they have, instead, chosen a promotional path: giving away two free months of ChatGPT Plus to any college student in the US or Canada, with the slogan “ChatGPT is here to help you through finals” (*ChatGPT plus Free for Students / ChatGPT*, 2025). Within weeks of the announcement of OpenAI’s free trial, xAI announced two free months of Super Grok for college students, and Google made the stunning announcement of offering over a year’s worth of free access to the premium version of Gemini, Google’s most powerful AI model (Woodward, 2025). Market share matters very much to these companies.

Indeed, GenAI companies realized early that the power users of their products were students, not working professionals or any other demographic group, and they are now targeting students around the world quite aggressively. Their hope, of course, is to win users and establish

loyalty. Popularity and profits are taking precedence over what is best for student learning. Mainstream AI companies are not directly marketing the concept of academic dishonesty, as are some marginal providers whose ads are likely prevalent on your students' social media feeds, but mainstream AI companies are facilitating such behaviors, which are obviously adverse to traditional learning. Where are teachers and professors in these conversations? How do we encourage GenAI companies to better align their marketing with their lofty public pronouncements about societal benefits?

Those are questions sloshing around in the tsunami. In a related concern, the competing announcements of free trials of premium GenAI models only for college students may also be a blow to equity and access. If GenAI is to become a universal tool to uplift humanity, as promised by the most eager optimists, then why grant free access only to college students? Giving the most powerful versions of AI only to people around the globe who have gained access to higher education is hardly an equitable framework for scaling this technology throughout society. A shrewd and calculated marketing move to hook users at a young age? Yes. A pathway toward empowering all people equally? No.

Imposed strategies related to rollout and marketing on the part of OpenAI, Google, and xAI will surely cause continued disruptions on our campuses. These are the ingredients of the tsunami, and the bots are too often separating human teachers from human learners. Faculty are left to forge temporary order from chaos and change. For example, without stability or standardization, we are now engaging many different learners: those who don't use GenAI, those who use the free versions of the tools, and those who have adopted trials or paid for access to premium features. The students are all sitting together in our classrooms. They have their laptops open, but they are difficult to tell apart and, of course, some of them are hiding their status.

The answer now and into the foreseeable future is open communication between teachers and students. If the bots are wedging themselves between us, we must name and nudge the bots. We must appeal to our students, rather than police them. It is an ethical path forward—and perhaps the only practical one. Faculty that rushed to adopt AI detection found the tools to be unreliable, and some innocent students have been caught up in this part of the tsunami. Perhaps detection will become possible in the future. In truth, the mainstream AI companies could enable watermarking tomorrow and give teachers this useful tool for transparency and tracking. Google has open-sourced watermarking through cryptography via SynthID, but few other AI developers have embraced such methods, which would contradict marketing efforts aimed at children and young adults. Even so, the promise of other detection schemes persist—stylometry, linguistic fingerprinting, and most recently process tracking in the form of Grammarly Authorship and Turnitin Clarity. Experts in academic integrity remain unconvinced about process tracking and its ability to curb AI misuse, noting recent advancements in GenAI Agents and Deep Research tools “have now made this strategy unreliable” (Bassett & Murdoch, 2025). Many of our colleagues want reliable policing tools, and some of our colleagues believe falsely that they can “spot” AI on their own, but, in truth, there is no reliable AI detection available right now.

We do not see the point in undertaking a detection arms race when it comes to GenAI. Rather, we plan to spend our energy guiding students by talking to them directly about their GenAI usage in practical terms. Viewing Generative AI as solely a cheating tool cripples any potential conversation about AI's impact on student learning and, more broadly, on their future lives as professionals and as human beings. We want to empower our students with sustainable lessons. We do not want to discourage our students with temporary punishments. Now is the time for talking with students about how this emerging technology will affect them and our shared society. So

many interesting, novel, and motivating questions are suddenly available! The conversations will not be possible unless faculty can trust students, and vice versa. As bell hooks observed, “the classroom, with all its limitations, remains a location of possibility” (*Teaching to Transgress*, 1994). Perhaps AI companies are setting new boundaries and creating new problems, but we, as teachers, can and must protect the all-important human relationships within our classrooms.

What strikes us with alarm is by not talking with students about their GenAI usage, academia is ceding that conversation directly to GenAI developers. Recent news about bizarre and troubling behavior in ChatGPT that produced “sycophantic” interactions should make us all consider why open dialogue with students about these new tools matters for many reasons, including public safety. In late April of 2025, OpenAI released a system update for ChatGPT that caused the model to engage in behavior that “skewed towards responses that were overly supportive but disingenuous.” To OpenAI’s credit, they quickly rolled back the update, noting “Sycophantic interactions can be uncomfortable, unsettling, and cause distress. We fell short and are working on getting it right” (*Sycophancy in GPT-4o: What Happened and What We’re Doing about It*, 2025). We cannot trust AI companies, especially as they fight with one another and race (sometimes recklessly) ahead. Part of our challenge is teaching students to be wary consumers and users of this astonishing technology. We want them analyzing and understanding the tsunami, even as they surf the coming waves.

When AI companies release technology to the public and scale it for free, they risk creating a series of second-order consequences that no one can foresee. Our students are heavily using this technology. Some faculty colleagues are lamenting this trend. Other faculty colleagues are embracing this trend. No one among us has any control. We are doing our best, sometimes in contradictory ways, to contend with this imposition from Silicon Valley. Looking ahead, we need better teacher training and coordination. We need more support from our institutional leaders. We need to move quickly beyond facile opinions and misinformed ideals. All of us, no matter our predilections, should talk openly about GenAI with our students. We should teach the debate, engaging the positive possibilities and the worrying risks. Doing so will help our students make the best possible decisions now and in the future, as they choose to use or avoid this technology in their daily digital lives.

The authors of the following issue are doing just this kind of nuanced thinking about Generative AI and education. They are deftly riding the wave. The six essays presented gathered here all grapple with how best to engage students and fellow educators. In “Building Critical AI Literacy: An Approach to Generative AI,” Kathryn Conrad and Sean Kamperman advocate for a critical AI literacy approach that examines power dynamics and ethical implications of AI technologies instead of merely teaching students how to use GenAI as a tool. Adrienne Carthon examines the complex relationship between AI technology and HBCU students in “The Stakes Are High. Are the Benefits Bountiful?: HBCU Students, AI, & the Power of Composition.” Kirkwood Adams and Maria G. Baker examine ChatGPT’s feedback on first-year writing essays in “Characterizing ChatGPT’s Feedback for FYW: Analyzing Feedback Responses to Inquiry-driven Essays.” Kathleen Kennedy and Anuj Gupta outline a thoughtful framework around AI and data acumen learning outcomes in “AI and Data Competencies: Scaffolding holistic AI literacy in Higher Education.” In “Enhancing Special Education with Generative AI: Suggestions from K-12 Special Education,” Amy Walter chronicles using ChatGPT in a middle school setting. Mila Zhu’s exploration of music generation tools reveals creative possibilities and hidden biases.

These scholars are calling upon us to think creatively and to engage our students fully within this new environment. As they demonstrate, we can teach AI literacy as a fundamental and

transdisciplinary skillset. We can do so by fostering spaces for critical examination, by promoting transparent usage, and by creating classroom cultures of curiosity and skeptical inquiry. We can pay attention to the bots, while putting the humans first. In doing so, we will help our students develop beyond blind consumerism, to become informed citizens capable of surfing even the biggest waves.

References

- Anthropic Education Report: How University Students Use Claude*. (2025). Anthropic.com. <https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude>
- Barnum, M., & Seetharaman, D. (2025, March 16). *There's a Good Chance Your Kid Uses AI to Cheat*. WSJ; The Wall Street Journal. <https://www.wsj.com/tech/ai/chatgpt-ai-cheating-students-97075d3c>
- Bassett, M. A., & Murdoch, K. (2025, February 23). *Ten persistent academic integrity myths*. Substack.com; Needed Now in Learning and Teaching. <https://needednowlt.substack.com/p/ten-persistent-academic-integrity>
- ChatGPT Plus free for students | ChatGPT*. (2025, April). ChatGPT. <https://chatgpt.com/students>
- Claude for Education | Partnering with Universities on Responsible AI*. (2025). Anthropic.com. <https://www.anthropic.com/education>
- hooks, bell. (1994). *Teaching to Transgress: Education as the Practice of Freedom*. New York: Routledge, p. 207.
- Lee, H.-P. H., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025, April). *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*. Microsoft Research. <https://www.microsoft.com/en-us/research/publication/the-impact-of-generative-ai-on-critical-thinking-self-reported-reductions-in-cognitive-effort-and-confidence-effects-from-a-survey-of-knowledge-workers/>
- Schmidt, Eric. (2025, May 16). "The AI Revolution is Underhyped" [Video]. YouTube. <https://www.youtube.com/watch?v=id4YRO7G0wE>
- Sycophancy in GPT-4o: What happened and what we're doing about it*. (2025). Openai.com. <https://openai.com/index/sycophancy-in-gpt-4o/>
- Woodward, J. (2025, April 17). *College students in the U.S. are now eligible for the best of Google AI—and 2 TB storage—for free*. Google. <https://blog.google/products/gemini/google-one-ai-premium-students-free/>

Building Critical AI Literacy: An Approach to Generative AI

Kathryn Conrad & Sean Kamperman

Abstract:

The rise of generative AI has been insufficiently met with opportunities for either educators or students to understand and evaluate GenAI systems. We believe our responsibility as educators includes critically examining these technologies as well as the rhetoric and assumptions that surround them. Our approach has been to provide, and champion, funded professional development opportunities and resources to enable the building of critical AI literacy for educators and, in turn, for students. Our essay explores critical AI literacy in the context of both AI literacy and critical literacy discourses, distinguishing between an AI literacy that entails a decontextualized, instrumentalist approach to teaching and learning about AI, and critical AI literacy that emphasizes the ways in which AI technologies are situated in larger systems of power. It then recounts the planning process and critical AI literacy components of the 2024 AI & Digital Literacy Institute at the University of Kansas, arguing for the value of sustained, collaborative learning opportunities across the educational spectrum.

Keywords: *critical AI literacy, AI ethics, critical digital pedagogy, K-12 professional development, generative AI, digital literacy, professional learning workshops*

Generative AI technologies have begun to disrupt education in a way not seen since the dial-up internet boom of the 1990s. Animating this disruption is not only—perhaps not even mostly—the technology itself, but the relentless juggernaut of hype and techno-determinist rhetoric that has accompanied its deployment. We believe that our responsibility as educators includes critically examining not only these technologies but also the rhetoric and assumptions that surround them. Helping students learn how to engage critically with generative AI is central to ensuring that they can make the most of their educational experiences and be prepared for responsible citizenship and rewarding lives and careers.

In order to help students toward these goals, teachers themselves must have the opportunity and support to fully understand the range of ethical and use issues with generative AI. We strongly believe empowering secondary and postsecondary teachers to explore AI issues is critical, not only to strengthen teacher morale and student learning, but for the health of our public discourse in an era of rampant misinformation. Yet, although many teachers nationwide are eager to understand the impact of AI technology in education, many American schools cannot offer the training and resources teachers need fully to learn about and explore this emerging technology (U.S. Department of Education, 2023).

Our approach to these challenges has been actively to provide, and champion, professional development opportunities and resources to enable the building of critical AI literacy for educators

and, in turn, for students. This past summer, we partnered with the National Humanities Center (NHC) to host a weeklong institute at the Hall Center for the Humanities on the University of Kansas, Lawrence, campus titled “AI & Digital Literacy: Toward an Inclusive and Empowering Teaching Practice” (AIDL ‘24). This institute, the first in a multi-year series, focused specifically on humanities educators in the greater Kansas City region, in an effort to build a regional network of teachers interested in understanding the impact of AI on humanities education. This essay describes the steps we have taken through this initiative to build critical AI literacy in our region.

Beyond Skills: Putting AI Literacy in Context

A focus on building critical AI literacy should not be confused with AI literacy initiatives whose primary goal is to teach AI skills. Since ChatGPT’s release in November 2022, a bewildering array of AI literacy workshops, webinars, trainings, and courses have come online, most purporting to teach participants how to use generative AI tools with barely a side helping of critical reflection on the technologies’ social and environmental impacts. Indeed, most “ethics” discussions, when present at all, have focused on “ethical use”—a phrase that serves as shorthand for questions about when to use AI and when and how to disclose that use—rather than on the ethical issues raised by the development, training, and deployment of the technology. In their scoping literature review of AI literacy initiatives in higher education, Laupichler et al (2022) reflect this prioritization of use over reflection, defining AI literacy as “the ability to understand, use, monitor, and critically reflect on AI applications without necessarily being able to develop AI models themselves.” Other formulations, such as Hibbert et al’s (2024) “A Framework for AI Literacy,” developed by academics and technologists at Barnard College, take a classic Vygotskian scaffolded approach that moves participants from understanding AI to using/applying it to analyzing/evaluating it to creating it. Under this configuration, critical thinking about AI is itself instrumentalized as a step toward using and building AI.

While useful, such approaches represent a particular set of values and priorities that proceed from techno-determinism: that is, the assumption that widespread adoption of generative AI technologies is inevitable, in the way cars became inevitable over the course of the 20th century and handheld computers are seemingly becoming inevitable now. Even approaches grounded more solidly in a critical or ethics-based framework take AI’s inevitability as their exigence: discussing MIT’s i2Learning modular AI literacy curriculum with *Education Week* reporter Alyson Klein, MIT professor Cynthia Breazeal says “‘the AI genie is out of the bottle [...] It’s the machine under everything. It’s critical for all students to have AI literacy if they are going to be using computers, or really, almost any type of technology’” (Klein, 2023; see also Hollands & Breazeal, 2024). After hearing this particular metaphor so many times, it’s easy to forget that the figure of the unbottled genie represents a theory of human desire that, while credible, does not fully capture how people relate to technology.

Our shift in emphasis from use to critical reflection, we should note, echoes earlier critical turns in literacy studies that attempt to reposition print-alphabetic literacy from a neat set of transferable skills to a culturally and sociohistorically embedded technological apparatus—a framework of writing technologies and the institutions that support them within which we humans operate (see for example Street, 1997; Kynard, 2013; Collin & Street, 2014; Rogers & O’Daniels, 2015; Graff, 2022). The shift toward a critical approach to AI acknowledges that the deployment of generative AI systems, like the deployment of mass media before it, serves particular interests

that may or may not align with our goals for our students (or, for that matter, their goals). Generative AI technologies—any technologies, for that matter—aren’t neutral or ‘mere tools’ as is often claimed, but are, as AI researcher Kate Crawford has described them, a “registry of power” (Crawford, 2021, 8). This focus on who and what generative AI serves, we believe, should be the starting point of discussions about whether, and perhaps how, to integrate generative AI into our teaching practice.

In this vein, it’s worth remembering that in the early 90s, when buzzwords like ‘computer-assisted teaching and learning’ and ‘networked classrooms’ were the rage, hedges to rhapsodic claims that personal computing and the internet would soon solve education’s equity issues were plentiful. In the field of rhetoric & composition, Faigley’s (1992) experiments with electronic discussion forums in the writing classroom led him to characterize networked electronic discourse as fragmented and somewhat allergic to truth; Barton (1994) described the negative effects of dominant discourses of technology on teaching and research; and Selfe & Selfe’s (1994) critique of the politics of personal computer interfaces directed scholars’ attention to the cultural biases embedded in graphical user interfaces, paving the way for critiques of web browsers and social media sites (cf Sorapure, Iglesby, & Yatchisin, 1998; Grabill, 2003; Arola, 2010 & 2017). Elsewhere, in more industry-focused fields such as human-computer interaction, scholars studied the implications of anthropomorphism for how humans understand, and interact with, machines (cf Nass, Steuer, & Tauber, 1994), prefiguring contemporary critical efforts to de-anthropomorphize AI (Bender et al, 2021; Shanahan, 2024). Such early work helped engender the critical, responsible, ethical, self-determined, liberatory, and care-oriented stances toward generative AI we seek to promote today.

Whatever one makes of the early 90s critiques, from the vantage of our current era of rampant mis- and disinformation, social fragmentation, political polarization, and screen addiction, the optimistic rhetoric of the early internet era looks to have been quite off. And so, when venture capitalist Ben Horowitz, discussing the virtues of techno-optimism, observes without a hint of irony that “the homeless of San Francisco have better access to information and knowledge than the President of the United States did in 1980” (Andreessen & Horowitz, 2023 October 25, 16:19-16:27), we know we’re alright rejecting claims that students *must* learn how to integrate AI into their workflows or end up on the streets.

Don’t Believe the Hype: Critical AI Literacy, Power, & Ethics

We thus distinguish between AI literacy and *critical* AI literacy along the following lines: whereas the term AI literacy often entails a decontextualized, instrumentalist approach to teaching and learning about AI, critical AI literacy emphasizes the ways in which AI technologies are situated in larger systems of power. This emphasis on power and sociohistorical context, we believe, helps learners weigh the risks and benefits of using generative AI technologies. To further articulate what we see as important distinctions between AI literacy and critical AI literacy, we draw from the work of one of our AIDL ’24 plenary speakers, Antonio Byrd. Byrd noted in his presentation that the discourse of a literacy crisis, applied to AI, echoes earlier alleged literacy crises that were, in the end, less about empowering teachers and students and more about imposing outside agendas on the classroom. In our current moment, tech companies like Google, Microsoft, Meta, Anthropic, and OpenAI, eagerly on the lookout for ways to deliver on promises of huge profits (or at least to make up their staggeringly high compute costs), have circled the education sector as a key target market for their products. Pursuit of financial return on investment, mingled with a

healthy dash of authentic techno-utopian zeal, has led to a weaponized crisis discourse that, in Byrd's words, "asks teachers to approach AI as a utilitarian technosolutionism" (Byrd, 2024, slide 7). Byrd's comments resonate with those of Roderic Crooks, who has explored how "digital divide" discourse also veils a techno-solutionist approach to problems of inequality that itself "sets the stage for future extraction" (Crooks, 2022; Crooks, 2024, 184).

A critical AI literacy approach similarly does not take AI industry hype claims at face value, nor does it embrace a "doomerist" outlook as its detractors sometimes allege. Rather, by "critical" we invoke the sense of the term as used in research and education: "critical" as "carefully evaluating" as well as "vital." As Conrad argues in "Blueprint for an AI Bill of Rights for Education" (2023), "critical AI literacy includes understanding how automated and/or generative systems work, the limitations to which they are subject, the affordances and opportunities they present, and the full range of known harms (environmental as well as social)." "Critical" in this case does not necessarily mean "fault finding," even as we acknowledge there are many faults to find: rather, it means approaching generative and other AI systems as objects of study, as is the practice in the emerging field of critical AI studies.¹ And it means that, instead of looking for ways to "reduce friction" in adoption and use of these technologies, as OpenAI's former CTO Mira Murati (OpenAI, 2024) and others pressing for adoption of generative AI would have it, we add productive friction back into the experience, turning "our attention to those seams we are seduced into not seeing" (Elam, 2023; see also Gonzales, 2023; Rosenzweig, 2024; Watkins, 2024; Conrad, 2024; Furze, 2024).

The "seams" to which Elam refers include the many known harms often mentioned in passing but smoothed over by AI "upskilling" workshops and webinars. The harms and risks entailed by current commercial generative AI models are numerous, including but not limited to the reproduction of cultural, ethnic, and social stereotypes (e.g. Abid, Farooqi, & Zou, 2021; Birhane et al, 2021; Gadiraju et al, 2023; Fergusson et al, 2023); the generation and circulation of inaccurate information, often referred to as "hallucinations" (Dahl et al, 2024; Goodlad and Stone, 2024); intellectual property infringement (Appel et al, 2023; Kupferschmid, 2023; Turkewitz, 2023); bias in AI detection software (Liang et al, 2023); exploitation of labor (Medina, 2023; Muldoon and Woo, 2023; Perrigo, 2023; Rowe, 2023), and harmful environmental impacts (Luccioni, Hernandez-Garcia, 2023; Mazucatto, 2024; Hogan, 2024).² As tools that reproduce the biases of their datasets and offer the most probable responses to a given prompt, reliance on them threatens to

1. Critical AI studies comprises a number of works including Kate Crawford's *Atlas of AI: Power, Politics, and the Planetary Cost of Artificial Intelligence* (2021) and Joy Buolomwini's *Unmasking AI* (2023), for instance, and inspired the founding of the journal *Critical AI*, the international group Critical Studies in AI and Education, and several special issues of academic journals. The journal *Critical AI* describes its approach as follows, and this description might serve as shorthand for the approach of critical AI studies: "Though rooted in critical methods from the humanities, social sciences, and arts, *Critical AI* works with technologists, scientists, economists, policy makers, health professionals, teachers, community organizers, legislators, lawyers, and entrepreneurs who share the understanding of interdisciplinary research as a powerful tool for building and implementing accountable technology in the public interest."

2. U Maine's Jon Ippolito has created a helpful acronym to articulate some these risks. IMPACT RISK stands for Infowar, Monopoly, Plagiarism/Privacy, Automated Labor, Climate Impact, Reality Distortion, Injustice, Stereotyping, and Knockoff Experiences.

undermine diversity of thought (Goodlad and Conrad, 2024, 41.) Reliance on LLMs also jeopardizes cognitive and educational development (Darvishi et al, 2023; León-Dominguez, 2024; Goodlad and Conrad, 2024; Riley & Bruno, 2024). Students seem to be aware that the adoption of generative AI is not uncomplicated and have reported a keen desire to develop an understanding of the ethics of AI use, according to a recent survey conducted by *Inside Higher Ed* (2024).

Yet pressure for teachers to adopt generative AI and look past those “seams” has been in place since at least the end of 2022 with the launch of ChatGPT, and the launch of platforms such as MagicSchool and Khanmigo have made it even easier for teachers to do so, even as these platforms run the risk of undermining teachers in the long term. As Conrad notes, “even as ed tech gurus are selling the ‘magic’ of time-saving to teachers, the quiet logic that animates the development of these tools is that what is being automated is actually what matters” (Conrad, 2024)—that is, the assessments and teaching aids, not the student-teacher relationships (until those are automated, too). Or, to shift to a seemingly opposite and yet related concern raised by many AIDL ‘24 participants: do tools like MagicSchool and Khanmigo save time so teachers can do *more* work? Are there cases where the pressure to produce ever more stuff that these tools create—more quizzes, brain teasers, role-playing games, songs—lead to overly scaffolded learning environments and diminished opportunities for independent learning? Or will the assumption that these tools are helping teachers—whether or not they actually are—lead administrators to assign more preparations, even when those are outside of the teachers’ area of expertise?

Taking the time to sort through questions like these, to understand the technology in both its larger context and its context of use, is essential; yet educators cannot be expected to add gaining critical AI literacy to their workloads without institutional support. Indeed, as Conrad notes, teachers have met the challenge of dealing with AI’s disruptions “in the wake of pandemic-driven pedagogical disruptions: shifts to (and from) online learning often in tandem with layoffs, austerity, and heightened workloads. From K-12 to elite research universities, educators have managed this technology-driven turbulence with minimal training, support, or guidance”; her “Blueprint for an AI Bill of Rights for Education” includes professional development as a key right for educators (Conrad, 2023). In the absence of institutional support, educators are at risk of finding that the only available professional development around generative AI is skewed toward AI adoption by the very companies who stand to gain most from it.

Lessons from the 2024 AI & Digital Literacy Institute

We are deeply committed, therefore, to helping to provide opportunities for professional development about AI that come from a place of respect not only for educational professionals’ time, but also their domain expertise as scholars and teachers. These opportunities recognize the significance of generative AI to the current teaching environment without proceeding from technodeterministic and technosolutionist assumptions of its inevitability, necessity, and value.

Having situated our own approach to critical AI literacy, we now describe how we attempted to put these ideas into practice at the 2024 AI & Digital Literacy Institute, which took place at the University of Kansas (KU) Hall Center for the Humanities on June 3-7, 2024. We should note that our framework for critical AI literacy continues to develop, and that some of the principles articulated above actually emerged *from* the discussions at the institute. Thus, rather than a blueprint for a critical AI literacy institute, what follows is a practice account of one way that critical AI literacy principles can find expression through cross-institutional dialogue and scaffolded critical making (Ratto, 2011) activities.

The Planning Process

Our process for planning the 2024 institute began early. On June 1, 2023, a combined team of KU and National Humanities Center (NHC) representatives convened a group of 35 teachers at the Hall Center for the Humanities for the “AI and Digital Literacy Educators’ Summit.” The goal of the summit was to gather insights from regional secondary education, community college, and university instructors who were interested in emerging AI technologies and digital literacy practices about what resources they needed to meet the new challenges posed by generative AI. To ensure space to facilitate conversation while ensuring representation from a diverse set of institutions, we accepted approximately 45 participants, of whom 35 were able to attend. Participants came to us from twelve high schools, seven 4-year colleges/universities, and three 2-year colleges/community colleges; they possessed varying levels of familiarity with generative AI technologies.

The one-day summit introduced participants to some of the ethical and use considerations surrounding emerging generative AI technologies by way of presentations, tool demonstrations, and a tool workshop. In a pre-event survey, many participants expressed interest in learning more about AI detection strategies; knowing the limitations and ethical risks of AI detection tools, we attempted to steer the day’s discussion instead towards strategies for revising assignments and assessments in ways that discourage AI use, by focusing our tool demonstrations on the technologies’ limitations and failure modes. We then asked participants to design an assignment in which students would summarize and analyze part of a novel (Ralph Ellison’s *Invisible Man*) considering three possible approaches: a “traditional” approach without regard for how students might use AI to complete the assignment; an “AI-aware” approach in which the assignment would be redesigned to anticipate student AI use (though not necessarily to encourage it); and an “AI-enriched/collaborative” approach in which AI use would be sanctioned or required as a core aspect of the assignment. We gave participants six learning objectives to consider as they weighed these possible approaches, corresponding with six core competencies: literary analysis, critical thinking, historical understanding, writing skills, appreciation for diverse experiences, and ability to engage peers in discussion.

While the conversations that followed this exercise were rich, we quickly realized a single afternoon would be hardly sufficient to engage educators in the kind of critical thinking, experimentation, tinkering, and dialogue needed to reconceptualize even one of their assignments; nor would a single day be enough to learn about the tools at the level necessary for a truly critical understanding of them. Reflecting our own goals for the demonstration activity, throughout the day and in surveys afterward, the educators consistently expressed a desire to develop curricular best practices rather than focusing solely on “punishing” AI use among students. Many expressed frustration that they had no dedicated time or space to develop policy, and indeed that they were not consulted about their institutions’ policies in the few instances where policies had been established; many schools did not—and still do not—have policies that clearly address AI. They also shared how valuable it was to connect across the secondary-postsecondary divide.

Building on that one-day summit, we proceeded with a similar goal for the weeklong 2024 AIDL institute: to empower Kansas City area humanities teachers to navigate the educational challenges and opportunities posed by generative AI. Three participants from the original summit served as a steering committee to help us ensure that our perspective as educators at a 4-year research university were not the only ones represented and that we were developing programming to address concerns shared by teachers across educational levels. We attempted to attract a diverse

range of participants and design a program that would put teachers into conversation with experts on ethical AI use and each other; introduce them to new tools and use cases; spur dialogue around ethical concerns, use considerations, and implementation challenges; and empower teachers to join efforts to craft ethical AI policies at their schools and districts. Of the 33 educators who applied, 27 were accepted and 23 were able to attend. The group included 20 teachers, two librarians and an IT specialist. The program featured scholars from KU, the University of Missouri-Kansas City, Duke University, and Harvard University; interactive demonstrations on foundation generative AI models such as ChatGPT and Gemini and two then-new AI teaching assistant tools (Khanmigo and MagicSchool); and collaborative design sessions where teachers created and presented on instructional resources and policy documents.

Critical AI Elements of the Institute

Multiple Opportunities for Processing & Dialogue Around Ethical Concerns

Each day opened with a plenary presentation and discussion, followed by a morning reflection. The ethics of generative AI use quickly became a recurring topic of conversation, partially by design (the resource Padlet we shared with participants contained links to many of the AI ethics articles cited in the above review of AI harms) and partially due to participant demand. We particularly wanted to ensure that conversations about ethics did not become quarantined to a small section of the agenda, as is typical of many of the AI literacy programs that treat it at all, or that they be limited to the question of “ethical use” that usually focuses on whether and how to cite AI, but rather that they percolate throughout the week as participants engaged around the urgent questions around labor, copyright, environmental impacts, and access that AI technologies pose.

Indeed, dialogue was central to our model from the earliest planning stages. In recognition of their status as experts in their subject domains and authorities on pedagogy in their local contexts, we wanted to give participants plenty of time and space to express their viewpoints and learn from each other, rather than trample them with an endless parade of lectures and workshops. This was the main reason for our insistence on an in-person institute: we sensed that a virtual institute, while more accessible, would not be conducive to the kind of relationship-building we hoped to accomplish. As our NHC partners point out, it’s mainly the hallway and “water cooler” conversations, not the lectures, that explain the efficacy of their teacher institute model, which we heavily drew from. We therefore structured a fair amount of flex time into the schedule to allow participants to chew on ideas from the readings, presentations, and large group discussions during chats over coffee or walks around campus. Additionally, we utilized Padlet to create opportunities to interact and share resources virtually, as well as a GroupMe chat.

These multiple avenues and opportunities to interact and take processing breaks proved especially necessary given the large amount of novel, technical, and often disturbing information we asked participants to digest. On the second day of the institute, we included an afternoon plenary on AI’s ethical and use considerations, where we covered a range of issues including environmental harms, labor exploitation, intellectual property issues, bias in datasets and output, problems with AI detection, epistemic justice, data sovereignty, “hallucinations,” and misinformation, and suggested guidelines to help teachers decide whether and how they might choose to use AI and how they might start to help students build critical AI literacy. The other presentations, in order, addressed text-generative AI’s affordances for literary production and analysis (Aarthi Vadde, Duke); AI and accessibility (James Basham, KU); critical digital pedagogy as antidote to

AI hype (Antonio Byrd, University of Missouri-Kansas City); and strategies for teaching analytical and expository writing in the age of AI (Jane Rosenzweig, Harvard). The blizzard of information and interpretations from the plenary presentations and constant stream of links and resources being added to the Padlet, which included pieces with often contradictory perspectives on AI, easily could have overwhelmed participants had we not structured in more flex time; even so, a few participants would indicate the desire for *more* processing time in their exit surveys.

One complaint we have heard is that the focus on ethical concerns frames AI in an overly negative light, discouraging teachers from exploring the technologies' potential benefits. To the extent that virtually all Web 2.0 platforms involve varying levels of moral compromise, it is easy to grow frustrated or overwhelmed when confronted with AI harms; if one is forced to weigh the risks and benefits of genAI, the implication is they should also be weighing the risks and benefits of social media and other ethically dubious technologies that many of us have simply learned to take for granted. Indeed, as advocates of critical AI literacy, we are in the rather uncomfortable position of pushing back against a particularly disruptive and ethically suspect technology in a landscape littered with ethically suspect technologies. One outcome of a critical conversation about AI is that it offers an opportunity to reconsider many technologies and their place in the classroom. Roderic Crooks notes, for instance, that “complex digital technologies place their users into particular relations of accumulation, extraction, and profit, frequently to the benefit of powerful corporations that operate these systems,” even as “computers maintain a durable appeal as vectors of improvement and racial uplift” (Crooks, 2022). And as Audrey Watters writes in *Teaching Machines: The History of Personalized Learning* (2021), the push to automate teaching and learning is over a century old; it is “a story of how education became a technocracy, and it’s a story about how education technology became big business” (9). A dialogue with educators at a variety of institutions with different populations and perspectives helps to situate a variety of digital technologies in both historical and local context, to ask questions about technological interventions that we have taken for granted, and, perhaps most importantly, give educators the confidence to take a seat at the table where decisions about technological deployment are made.

Instructional Resource Development

Afternoons at the institute consisted mainly of critical making activities: workshops, interactive tool demonstrations, and group work time. Ratto (2011) defines critical making as “a mode of materially productive engagement that is intended to bridge the gap between creative physical and conceptual exploration” (252). The term “signals a desire to theoretically and pragmatically connect two modes of engagement with the world that are often held separate—critical thinking, typically understood as conceptually and linguistically based, and physical ‘making,’ goal-based material work” (253).” While we did not explicitly use the term “critical making” to conceptualize the work we asked participants to do as they experimented with AI tools, probing their limitations and assessing their affordances for various writing and research tasks, the act of creating AI instructional resources (and policy resources, as we describe later) undeniably led to startling insights about the technology that would not have surfaced from discussion alone.

The first day’s workshop riffed on the activity from the Summit we describe above, except this time, instead of asking participants to “AI-proof” or integrate AI into a traditional writing assignment, we asked them to approach an assignment as though they were the students. Participants were put into groups and instructed to use a foundational generative AI tool or tools to complete a portion of an essay on climate change: background research, an intro paragraph, thesis,

topic paragraph, etc. This activity revealed to many participants the vulnerabilities of traditional writing assignments to disruption by AI tools while also surfacing the tools' surprising limitations. Participants found, for instance, that so-called "research" was often comprised of fabricated links to non-existent sources, thesis statements and paragraphs generated in a particular style or "voice" sounded either bland or stereotypical, and arguments lacked both originality and meaningful connection to local contexts.

The following day, having experimented with foundation models, participants were invited to probe the affordances of Khanmigo and MagicSchool as part of an unstructured workshop. These tools, which at the time had just been widely released, are now well-known to educators (MagicSchool's website, at time of writing, claims over 5,000 schools and districts as "partners"). Not being deeply familiar with the tools ourselves, we were curious to see how participants would react to their "teacher-friendly" look and feel (and indeed, we learned that a few participants' schools had already signed up for institutional licenses). While many participants said they saw the potential usefulness of being able to easily generate lesson plans, discussion prompts, and quiz questions, this workshop prompted a fascinating discussion of the potential implications of these tools for teacher workloads as well as for student learning. By automating the "busy work" of lesson planning, would these platforms really free up teachers to focus on building relationships with their students and to provide increased support to struggling learners? Or would they simply facilitate further budget consolidations at already strapped schools, increasing the workloads of the teachers fortunate enough to keep their jobs (see e.g. Williamson, 2024)? Is the particular vision of student-centered education Khanmigo and MagicSchool profess—one where teachers can instantly, as if by magic, create any kind of learning aid they think their students might need, and where students enjoy 24/7 access to AI tutors (leaving out the issue of the inevitable inaccuracies of LLMs, of course)—conducive to helping students develop valuable problem-solving and socioemotional skills? To take one example, is it always better to make more slide decks and handouts, or might it be better in some classes for students to learn to take better notes? The answers to these questions may of course differ from school to school, student to student, and subject to subject. The point is that, by taking time to experiment with and discuss these tools in a collegial, cross-institutional setting, we arrived at critical questions we might not have otherwise.

Having thus prompted participants to consider a critical approach to the tech, we put the participants in teams and set them loose to create instructional resources that either foregrounded generative AI or used generative AI, with the group members offering reflections on the activities and their benefits and limitations. While we attempted to group participants by grade level, some groups were more vertically integrated, which led to interesting comparisons and adaptations of the resources to suit multiple audiences. Each resource was shared on the AIDL Padlet and will be posted to the NHC's website, where they will be available to teachers around the country. These included the following (with recommended audience in parentheses):

Group 1: The Passive Voice: 1st Year Writing Course Grammar Lesson (Community College)

This lesson involves the use of several free AI tools for educators and students offered by Khanmigo and MagicSchool to help generate learning objectives and activities for teaching strategic use of passive voice. These activities, including AI-generated skits, not only provide an engaging way to understand the passive voice, but when shared with students transparently as AI-generated material, also provide opportunities to teach critical AI literacy. In particular, the lesson presents an opportunity to engage interesting and important questions about genAI's status as a

“knower”: members of this group found that Magic School’s AI chat assistant (a unicorn named “Raina”) would occasionally misidentify prose passages as containing passive voice when they did not.

Group 2: AI Biography (High School and Up)

This lesson is explicitly focused on building AI literacy for high school or college students, with a focus on meeting the International Society for Technology in Education (ISTE) Standard of helping students learn to “evaluate the accuracy, perspective, credibility and relevance of information, media, data or other resources.” Students would be guided to use free AI tools to generate a biography, with the teacher suggesting different options and tools and students sharing and reflecting on the outputs and errors as one way to help them understand the possibilities as well as the limitations of generative AI models.

Group 3: Using AI to Parse Shakespeare (Middle and High School)

This resource uses several AI models and different prompt layers to help middle and high school students parse challenging Shakespearean texts. The intent was to provide ways for students to understand Shakespeare and linguistic differences by using paraphrases generated by AI models in different dialects. The resource in particular utilized Latimer, an app built on OpenAI but also “built to more accurately reflect the experience, culture and history of Black and brown people” in part through its training on a secondary curated dataset. Use of apps like Latimer can help students probe questions about bias in large language models.

Group 4: Create & Analyze a Song (High School and Up)

This lesson begins with students using MagicSchool’s AI song generator to create a song in the style of a particular singer and then analyzing that song to see how successful it was at capturing the artist’s style. Students would be encouraged to help generate a rubric to evaluate whether the song achieves its stated goal, which allows students not only to reflect on elements of poetry studied using traditional texts but also to reflect on the tool itself and its strengths and limitations. The lesson goals include building the skills of close reading, analysis, and synthesis; understanding genre, literary devices, and style; and building critical AI literacy.

Group 5: Friends? Enemies? Frenemies? Exploring Rhetoric & Subject Positionality With/Against AI Chatbots (College)

In this lesson, students build on literary criticism and theory to design and program chatbots to represent a variety of hypothetical relationships (e.g. friend, enemy, etc). Students would be guided to evaluate and adjust their chatbot's discourse parameters to enhance the quality of their interaction with it and in so doing, to explore rhetorical assumptions embedded in the technology. The primary graded element of the activity would be a presentation in which students formulate an argument about the opportunities, limitations, biases, etc. of LLMs and provide and analyze specific examples from their chatbot transcripts. The group said they were inspired to create this lesson when they began to notice, and question, the obsequious politeness of ChatGPT and other student- and customer-facing chatbots. How did these friendly bots (and the corporations behind

them) position their human interlocutors? What do these types of discourse features reveal about the deeper human-technology relations and interactions chatbot interfaces are designed to promote?

Policy Crafting Sessions

During our preparation for the Institute, our steering committee articulated their interest in having at least one session focused on policy. Our decision to build this in, and our decision to focus one of the group work sessions on policy creation, ended up anticipating one of the refrains over the course of the week from our participants: the need to prepare educators to be at the table when policy and technology decisions were made, from the classroom to the district and regional levels.

We began the policy conversation by asking participants to bring in any current policy from their institutions, which in turn prompted a larger discussion wherein participants articulated what they believed that success in policymaking would look like. We shared “Cross-Campus Approaches to Building a Generative AI Policy” (Brandon et al, 2023). Then we let participants choose groups based on whether they wanted to create policy for their classroom, department, institution/school, or district. Groups were asked to consider what key areas their policy should address, as well as the following questions raised by Brandon et al:

- **Policy Audience:** Whom is this policy for? Is it for the entire institution, faculty, students, staff, departments, third-party vendors, or others?
- **Policy Timeline:** What is the timeline for implementation? Should that timeline include a review and update cycle after initial implementation?
- **Policy Tools:** What counts as AI for this policy? Is it focused on all AI or only generative AI? Is it focused on all generative AI or just text-generating AI?
- **Academic Integrity Guidelines:** If the policy is related to student or faculty work, what are the integrity requirements to make sure academic integrity is upheld?
- **Acceptable Use:** If AI usage is acceptable, are there any limitations on the amount of usage (e.g., a certain percentage must be individually generated in certain contexts) or purposes for which generative AI may be used?
- **Transparency:** What practices are in place for communicating its usage throughout the institution?
- **Security and Legal Considerations:** What concerns need to be addressed concerning privacy, intellectual property, and proprietary knowledge around using external or enterprise generative AI tools? Does endorsing the use of generative AI conflict with any laws such as the General Data Protection Regulation?
- **Ethical Considerations:** What concerns or responsibilities does the institution have explicitly or implicitly within its mission that conflict with the environmental, human-exploitation, and bias issues related to generative AI?
- **Institutional Resources:** Which areas of the institution will be committed to supporting, responding to, and implementing uses of generative AI?
- **Processes for Policy Violation:** If the policy is violated, what are the steps for identifying and addressing it? (adapted from Brandon et al, 2023).

Two of our participants used their group policy project as a starting point to write a policy blueprint for a local school district, and several articulated their plans to take the documents they had created, and the policy conversation guidance we had provided, back to their own institutions to animate policy conversations. These documents ranged from statements of principles to classroom posters to instructions for facilitating student discussions about AI ethics and best practices in the classroom.

Conclusions

For us, the NHC teacher institute model, with its emphasis on participant dialogue, intensive resource development, and sustainable relationships, has proven to be an excellent foundation for further refining our approach to critical AI literacy. Educators at all levels need time to talk to each other, learn about how LLMs work, absorb expert opinions and recommendations, become familiar with tools, update teaching resources and create new ones, think through policy, and build alliances—just as students require the flexibility in the classroom environment to do the same. All of these activities, we believe, are indispensable to developing a more informed, socially aware pedagogical praxis, defined by technical communication scholar Natasha Jones (2016) as “an essential, not-to-be-separated pairing of reflection and action” (p. 224). A sound pedagogical praxis can help us form ethically grounded justifications for our choices whether or not to use AI or have our students use it, protecting against further harms and violations of rights.

While one-day AI workshops can help foster dialogue, prompt reflection, and teach skills in a limited way, weeklong in-person experiences have the power to build networks and foster the committed, sustainable action necessary for praxis to take hold and for institutional transformation to occur. 8- to 16-week courses, we’d note, afford similar opportunities for students to slow down their thinking; in principle, the activities and discussions described above could all be modified for a college course. Flipping the classroom and having students create assignments and draft classroom policies could be an engaging way to teach responsible and ethical use, while helping students take more responsibility for their learning.

Our experiences with AIDL ‘24 have also shown us the value of having conversations about AI that cut across the traditional divide between secondary and postsecondary education, as well as divides between community colleges and four-year institutions. At the college level, the mysteriousness of our students’ educational backgrounds is a fact often taken for granted, at least from the instructor’s point of view; we treat them more or less as blank slates and subject them to a curriculum of uniform expectations regardless of their prior experiences. From conversations with middle and high school teachers, we’ve learned that this uncertainty pervades the other side of the secondary/postsecondary divide as well. The ad hoc fixes to this systemic issue include loads of administrative support and some excellent bridge programs, as well as increasingly refined first-year experience curricula. Many first-year writing programs, including KU’s, include lessons and assignments in the first few weeks of class designed to excavate students’ literacy habits, skills, and experiences, to give instructors some kind of baseline. Curiously, we’ve found that conversations between secondary and postsecondary educators rarely seem to enter the mix of solutions in any formal way, partially as a result of a professional development funding landscape that keeps high school and college teachers in their separate bubbles (the NEH, for example, runs separate grant programs for high school teacher institutes and programs targeted at college-level educators; our 2025 version of the institute, originally funded by the NEH until DOGE’s cuts, appeared to be the only NEH-funded program that explicitly bridged that divide). The shockwaves generated by

genAI have awoken many to the learning barriers these traditional silos create. Learning how middle and high school teachers in our region are teaching about AI and the shifting institutional constraints they are operating under has expanded our thinking about what a cross-level critical AI curriculum can, and should, entail. These conversations are valuable even—or especially—when they challenge assumptions across the secondary/postsecondary divide: just as “friction” served as a useful metaphor for critical AI literacy pedagogy, so too does it serve us as a way to describe where we as educators can stop and identify moments of productive discussion and exploration. We further believe that by encouraging cross-level conversations, we sow the seeds necessary for system-wide interventions, whether those are policy conversations at a district or international level (such as the recommendations on AI to the UN from the Special Rapporteur for Education for which Conrad served as an advisor).

Amongst the range of other approaches to generative AI, critical digital pedagogy as theorized by Byrd, Stommel, and others, invite both students and teachers to interrogate the power relations inherent in both physical and digital learning environments as a means to becoming “full agents of their own learning” (Stommel, 2014) and, ultimately, toward liberation. A critical AI literacy approach sits comfortably alongside this approach, while arguably retaining a greater emphasis on the “what” and the “how to” of genAI—not in order to teach skills, but to further intertwine thinking critically about AI with the doing of AI. The critical AI literacy model we have chosen thus requires that educators develop a solid technical understanding of how LLMs work, without which ethical choices about whether and how to use AI become impossible. Building basic technical literacy is something we plan to incorporate more intentionally in future institutes; without that technical literacy, it is easier to fall prey to the kind of hype that accompanies the pressure to adopt AI tools.

Our experience in this institute, and the enthusiastic response to it from participants in the post-institute surveys, ultimately suggest to us the need for more such opportunities: opportunities for conversation, productive friction, and critical making. These opportunities should be provided with respect for and valuing of educators’ time through professional development credit and funding. They should also be designed with respect for educators across learning levels by treating all participants as colleagues with perspectives and experiences that matter. Our own efforts to build critical AI literacy through teacher institutes have yielded unexpected gifts, not only helping us better understand this technology that is shaking up our classrooms, but also connecting us in real and vital ways to colleagues across the educational spectrum—a satisfyingly human outcome in an increasingly dehumanized educational landscape.

References

- Abid, A., Farooqi, M. & Zou, J. (2021) Persistent anti-Muslim bias in large language models. *ArXiv*. <https://doi.org/10.48550/arXiv.2101.05783>
- Alenichev, A., Kingori, P., & Peeters Grietens, K. (2023, August 9) Reflections before the storm: the AI reproduction of biased imagery in global health visuals. *The Lancet Global Health*. [https://doi.org/10.1016/S2214-109X\(23\)00329-7](https://doi.org/10.1016/S2214-109X(23)00329-7)
- Andreessen, M. & Horowitz, B. (Host). (2023, October 25). The techno-optimist manifesto (No. 747) [Audio podcast episode]. In *a16z Podcast*. Andreessen Horowitz. <https://a16z.com/podcasts/a16z-podcast/>
- Appel, G., Neelbauer, J., & Schweidel, D. A. (2023, April 7). Generative AI has an intellectual property problem. *Harvard Business Review*. <https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>

- Arola, K. L. (2010). The design of Web 2.0: The rise of the template, the fall of design. *Computers and Composition*, 27(1), 4-14.
- Arola, K. L. (2017). Indigenous interfaces. In D. Walls & S. Vie (Eds.), *Social writing/social media: Publics, presentations, and pedagogies*, (pp. 209-224). The WAC Clearinghouse; University Press of Colorado.
- Bali, M. (2023, April 1). What I mean when I say critical AI literacy. Reflecting Allowed [blog]. <https://blog.mahabali.me/educational-technology-2/what-i-mean-when-i-say-critical-ai-literacy/>
- Barton, E. (1994). Interpreting the discourses of technology. In C. Selfe & S. Hilligoss (Eds.), *Literacy and computers: the complications of teaching and learning with technology* (pp. 56-75). Modern Language Association of America.
- Beck, S. W., & Levine, S. R. (2023). Backtalk: ChatGPT: A powerful technology tool for writing instruction. *Phi Delta Kappan*, 105(1), 66-67. <https://doi.org/10.1177/00317217231197487>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).
- Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., & Caliskan, A. (2023, June). Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. *FACCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (1493–1504) <https://doi.org/10.1145/3593013.3594095>
- Birhane, A., Uday Prabhu, V., & Kahembwe, E. (2021, October). Multimodal datasets: misogyny, pornography, and malignant stereotypes. *ArXiv*. <https://doi.org/10.48550/arXiv.2110.01963>
- Brandon, E., Eaton, L., Gavin, D., & Papini, A. (2023, Dec. 12). Cross-campus approaches to building a generative AI policy. *Educause Review*. <https://er.educause.edu/articles/2023/12/cross-campus-approaches-to-building-a-generative-ai-policy>
- Buolomwini, J. (2023) *Unmasking AI*. Random House.
- Byrd, A. (2024, June). Using critical digital pedagogy to promote liberatory teaching during the ‘AI literacy crisis’ [slideshow].
- Collin, R., & Street, B. V. (2014). Ideology and Interaction: Debating Determinisms in Literacy Studies. *Reading Research Quarterly*, 49(3), 351–359. <https://doi.org/10.1002/rrq.75>
- Conrad, K. (2023). Blueprint for an AI Bill of Rights for education. *Critical AI* 2.1. <https://criticalai.org/2023/07/17/a-blueprint-for-an-ai-bill-of-rights-for-education-kathryn-conrad/>
- Conrad, K. (2024). Friction v. “Magic”: the lure of AI time-saving tools for teachers. *Pandora’s Bot* [Substack blog]. <https://kconrad.substack.com/p/friction-v-magic>
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale.
- Crooks, R. (2022, Jan. 26). Toward people’s community control of technology: Race, access, and education. *Just Tech*. Social Science Research Council. DOI: <https://doi.org/10.35650/JT.3015.d.2022>
- Crooks, R. (2024). *Access Is capture: How Edtech reproduces racial inequality*. University of California Press.
- Dahl, M., Magesh, V., Suzgun, M., & Ho, D. E. (2024, Jan. 11) Large legal fictions: Profiling legal hallucinations in large language models. *ArXiv*. <https://doi.org/10.48550/arXiv.2401.01301>

- Darvishi, A., Khosravib, H., Sadiqb, S., Gašević, D., & Siemens, G. (2023, November 30). Impact of AI on student agency. *Computers & Education*. <https://doi.org/10.1016/j.compedu.2023.104967>
- Elam, M. (2023). Poetry will not optimize; or, what is literature to AI? *American Literature* (2023) 95 (2): 281–303. <https://doi.org/10.1215/00029831-10575077>
- Estrada, D. (2023). Teaching insights: How to teach AI to students (AI Ethics and @NJIT Audit Project). *Critical AI* blog. <https://criticalai.org/2023/05/30/teaching-insights-how-to-teach-ai-to-students-ai-ethics-and-njit-audit-project-dr-daniel-estrada/>
- Faigley, Lester, F., Lester. (1992). *Fragments of rationality : postmodernity and the subject of composition*. University of Pittsburgh Press.
- Fergusson, G., Schroeder, C., Winters, B., & Zhou, E., eds. (2023, May) *Generating harms: Generative AI's impact and paths forward*. Electronic Privacy Information Center. <https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf>
- Flaherty, C. (2024) Survey: How AI Is impacting students' career choices. *Inside Higher Ed*. (10 Jan 24) <https://www.insidehighered.com/news/student-success/life-after-college/2024/01/10/survey-college-students-thoughts-ai-and-careers>
- Furze, L. (2024) Why writing needs good friction. *Leon Furze* [blog]. <https://leonfurze.com/2024/07/04/why-writing-needs-good-friction/>
- Gadiraju, V., Kane, S., Dev, S., Taylor, A., Wang, D., Denton, E., & Brewer, R. (2023, June) "I wouldn't say offensive but...": Disability-centered perspectives on Large Language Models. FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (205-215). <https://doi.org/10.1145/3593013.3593989>
- Gonzales, M. R. (2023, Sept. 13). Artificial mediocrity: The hazard of AI in education. *The Public Discourse*. <https://www.thepublicdiscourse.com/2023/09/90834/>
- Goodlad, L. M. E. & Conrad, K. (2024, March 14). Teaching critical AI literacies. Policy Insights: AI and Digital Inequities. NORRAG. <https://resources.norrag.org/resource/view/845/468>
- Goodlad, L. M. E. & Stone, M. (2024). Beyond Chatbot-K: On Large Language Models, “Generative AI,” and rise of chatbots—an introduction. *Critical AI* 2 (1). <https://doi.org/10.1215/2834703X-11205147>
- Grabill, J. T. (2003). On divides and interfaces: Access, class, and computers. *Computers and composition*, 20(4), 455-472.
- Graff, H. J. (2022). *Searching for literacy: The social and intellectual origins of literacy studies*. Palgrave Macmillan.
- Gupta, R. (2024, January 1). LAION and the challenges of preventing AI-Generated CSAM. *TechPolicy.Press*. <https://www.techpolicy.press/laion-and-the-challenges-of-preventing-ai-generated-csam/>
- Hibbert, M., Altman, E., Shippen, T., & Wright, M. (2024). A framework for AI literacy. EDU-CAUSE Review. Retrieved September 30, 2024, from <https://er.educause.edu/articles/2024/6/a-framework-for-ai-literacy>
- Hogan, M. (2024). The fumes of AI. *Critical AI* 2 (1). <https://doi.org/10.1215/2834703X-11205231>
- Hollands, F., & Breazeal, C. (2024). Establishing AI Literacy before Adopting AI. *The Science Teacher*, 91(2), 35–42. <https://doi.org/10.1080/00368555.2024.2308316>
- Ippolito, J. (2024). Impact risks: An acronym for AI downsides. U Maine. https://nmdprojects.net/learnwithai_www/impactrisk/

- Jiang, H. H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., Hanna, A., Flowers, J., & Gebru, T. (2023) AI art and its impact on artists. AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, pp. 363–374. <https://doi.org/10.1145/3600211.3604681>
- Klein, A. (2023, May 10). AI Literacy, Explained. *Education Week*. <https://www.edweek.org/technology/ai-literacy-explained/2023/05>
- Kupferschmid, K. (2023, June 12). Copyright alliance, AI accountability policy request for comment, Docket No. 230407–0093. <https://copyrightalliance.org/wp-content/uploads/2023/06/NTIA-AI-Comments-FINAL.pdf>
- Kynard, C. (2013). *Vernacular insurrections: Race, black protest, and the new century in composition-literacies studies*. State University of New York Press.
- Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, 100101. <https://doi.org/10.1016/j.caeai.2022.100101>
- León-Domínguez, U. (2024). Potential cognitive risks of generative transformer-based AI chatbots on higher order executive functions. *Neuropsychology*, 38(4), 293–308. <https://doi.org/10.1037/neu0000948>
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023, April 6) GPT detectors are biased against non-native English writers. *ArXiv*. <https://arxiv.org/abs/2304.02819>
- Luccioni, A. S. & Hernandez-Garcia, A. (2023). Counting carbon: A survey of factors influencing the emissions of machine learning. *ArXiv*. <https://doi.org/10.48550/arXiv.2302.08476>
- MagicSchool. (2024). <https://www.magicschool.ai/> Accessed Sept. 9, 2024.
- Marcus, G. & Southen, R. (2024, Jan 6). Generative AI has a visual plagiarism problem. *IEEE Spectrum*. <https://spectrum.ieee.org/midjourney-copyright>
- Mazzucato, M. (2024, May 31). The ugly truth behind ChatGPT: AI is guzzling resources at planet-eating rates. *The Guardian*. https://www.theguardian.com/commentisfree/article/2024/may/30/ugly-truth-ai-chatgpt-guzzling-resources-environment?CMP=fb_a-technology_b-gdntech
- Medina, P. (2023, Jun. 24). The Venezuelan ghost workers who are feeding artificial intelligence. *El País*. <https://english.elpais.com/international/2023-06-24/the-venezuelan-ghost-workers-who-are-feeding-artificial-intelligence.html>
- Muldoon, J. & Wu, B. A. (2023, Dec. 15). Artificial Intelligence in the Colonial Matrix of Power. *Philosophy & Technology* 36, article 80. <https://link.springer.com/article/10.1007/s13347-023-00687-8?fbclid>
- Nass, C., Steuer, J., & Tauber, E. R. (1994, April). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78). <https://doi.org/10.1145/191666.191703>
- OpenAI (2024). Introducing GPT-4o. [video] <https://www.youtube.com/watch?v=DQacCB9tDaw>
- Perrigo, B. (2023, Jan. 18). OpenAI Used Kenyan workers on less than \$2 per hour to make ChatGPT less toxic. *Time*. <https://time.com/6247678/openai-chatgpt-kenya-workers/>
- Riley, B., & Bruno, P. (2024). Education hazards of generative AI. *Cognitive Resonance*. <https://cognitiveresonance.net/EducationHazardsofGenerativeAI.pdf>
- Rogers, R. & O'Daniels, K. (2015). Critical literacy education: A kaleidoscopic view of the field. In J. Roswell & K. Pahl (Eds.), *The Routledge Handbook of Literacy Studies* (1st ed.), pp. 62-78. Routledge. <https://doi.org/10.4324/9781315717647>

- Rosenzweig, J. (2024) When friction is the point. *Writing Hacks* [Substack blog]. <https://writinghacks.substack.com/p/when-the-friction-is-the-point>
- Rosenzweig, J. (2023) Teaching Insights: What happens when a novice writer asks ChatGPT for editing advice? *Critical AI* blog. <https://criticalai.org/2023/09/21/teaching-insights-what-happens-when-a-novice-writer-asks-chatgpt-for-editing-advice-dr-jane-rosenzweig/>
- Rowe, N. (2023, Oct. 16). Millions of workers are training AI models for pennies. <https://www.wired.com/story/millions-of-workers-are-training-ai-models-for-pennies/>
- Selfe, C. L., & Selfe, R. J. (1994). The Politics of the Interface: Power and Its Exercise in Electronic Contact Zones. *College Composition and Communication*, 45(4), 480–504. <https://doi.org/10.2307/358761>
- Shanahan, M. (2024). Talking about large language models. *Communications of the ACM*, 67(2), 68–79.
- Sorapure, M., Inglesby, P., & Yatchisin, G. (1998). Web literacy: Challenges and opportunities for research in a new medium. *Computers and Composition*, 15(3), 409–424.
- Stommel, J., Friend, C., & Morris, S. M. (2020) *Critical Digital Pedagogy: A Collection*. Hybrid Pedagogy Inc.
- Stommel, J. (2014). Critical digital pedagogy: A definition. Hybrid Pedagogy [web]. <https://hybridpedagogy.org/critical-digital-pedagogy-definition/>
- Street, B. (1997). The Implications of the ‘New Literacy Studies’ for Literacy Education. *English in Education*, 31(3), 45–59. <https://doi.org/10.1111/j.1754-8845.1997.tb00133.x>
- Thiel, D. (2023, 20 Dec.) Investigation finds AI image generation models trained on child abuse. *Stanford Cyber Policy Blog*. <https://cyber.fsi.stanford.edu/news/investigation-finds-ai-image-generation-models-trained-child-abuse>
- Turkewitz, N. (2023, February 22) “The fair use tango: A dangerous dance with [re]generative AI models.” Neil Turkewitz, *Medium* [blog]. https://medium.com/@nturkewitz_56674/the-fair-use-tango-a-dangerous-dance-with-re-generative-ai-models-f045b4d4196e
- U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning*. <https://www2.ed.gov/documents/ai-report/ai-report.pdf>
- Watkins, M. (2024). We need to reclaim slowness. *Rhetorica* [Substack blog] <https://marcwatkins.substack.com/p/we-need-to-reclaim-slowness>
- Watters, A. (2021). *Teaching machines: The history of personalized learning*. MIT Press.
- Williamson, B. (2024). AI in education is a public problem. *Code Acts in Education* [blog] <https://codeactsineducation.wordpress.com/2024/02/22/ai-in-education-is-a-public-problem/>

Characterizing ChatGPT's Feedback for FYW: Analyzing Feedback Responses to Inquiry-driven Essays

Kirkwood Adams & Maria G. Baker

Abstract

In response to 1) studies finding that essay feedback generated by ChatGPT might be useful for student writers and 2) studies observing ChatGPT's tendency to adhere to narrow genre definitions when producing writing, our study seeks to examine whether ChatGPT can provide useful feedback in a first-year writing learning environment that targets a more nuanced genre definition of the academic essay as a form of inquiry. Our study employed three sequentially constructed feedback request prompts to assess how ChatGPT would go about generating outputs when asked to respond to a set of 50 published student essays that successfully achieve the goals of Columbia University's FYW class. The use of exemplary essays as input data revealed a durability in ChatGPT's feedback priorities under the influence of more robust prompting strategies. We found that in two-part responses, ChatGPT provided a descriptive as well as a prescriptive portion of feedback, the former largely affirmative, the latter critical with a focus on revision suggestions that guide toward narrow genre markers. The descriptive feedback mode seemed more adaptable to our capacious definition of genre, and the model responded less critically to exemplary essays when in descriptive mode, suggesting that further study of descriptive feedback capacities might be helpful in developing ChatGPT's potential to give useful feedback to students in FYW courses. However, in two brief stress tests of our results, we found cause for amending our findings.

Keywords: *first-year writing, generative artificial intelligence, feedback, genre, machine intelligence*

Introduction

The project of interacting with LLMs to supplement students' learning can be succinctly illustrated by "Assigning AI: Seven Approaches for Students, with Prompts", a primer by Mollick & Mollick (2023), co-directors of the Generative AI Lab at the Wharton School. Amongst other applications, Mollick and Mollick envision the theoretical potential of genAI systems to provide feedback on student writing. Much of Mollick and Mollick's argument hinges upon the major selling points which warranted the release of now commercially available genAI systems in the first place: claims that LLM chat conversations can be personalized, an emphasis on the speed of the conversational exchange of user input and system output, and the scalable application of these systems for use by all/any student(s), who each can have their own fast, customized conversation. Although Mollick and Mollick do not teach writing themselves, they offer far-reaching suggestions

for it can be practiced and learned with genAI: purposefully prompting genAI to role-play as a “mentor” so that these huge, universally-oriented systems made by OpenAI or Google can provide students with “balanced and on-going feedback” (p. 3). Elsewhere, quantitative researchers have begun to systematically test such a theoretical proposition, assessing ChatGPT, the LLM which has become a household name, to test its potential for giving feedback on writing in specific classroom contexts. Researchers concluded that ChatGPT demonstrated the capacity to give high-quality feedback for high school students studying history (Steiss et al., 2024), for graduate-level life science students (Banihashem et al., 2024), & postgraduate students studying data science (Dai et al., 2023).

Given the reasonable proposition of a feedback use-case for genAI and the preliminary conclusions that high-quality feedback on student writing is feasible in specific teaching and learning contexts, other educators may wonder whether commercially available genAI systems would provide useful responses to their own students’ writing. Having undertaken the study of genAI systems ourselves, as teachers in Columbia’s Undergraduate Writing Program, we wonder specifically whether receiving feedback from ChatGPT would be effective for students taking a First Year Writing course. Bringing this feedback question to a FYW context offers an interesting test case. In FYW courses, student writing is not an assessment instrument for another discipline’s learning goals, for instance, the successful uptake or application of knowledge specific to that discipline. Rather, developing capacities as writers is the overarching learning goal of FYW. We would first need to assess whether ChatGPT’s feedback would align with and support the learning goals of FYW before incorporating it. One approach to assessing the quality of ChatGPT’s feedback hinges on whether this system shares a fundamental understanding of the essay as a genre with us.

Genre & Its Role in Feedback

When prompted to define the academic essay as a genre, an exercise we often do with our FYW students, ChatGPT consistently offers a highly formalized set of conventions as constitutive, as illustrated in this sample output:

An academic essay is a specific genre of writing that aims to inform, argue, or analyze a particular topic or question within a scholarly framework. This genre is characterized by a clear, focused thesis or central argument, supported by evidence and critical reasoning. It adheres to formal conventions of structure, style, and citation, making it a disciplined and methodical form of communication. (Open AI)

In ChatGPT’s verbose way, this output continues, further elaborating on each of these elements and explaining how academic essays feature introductions (whose primary purpose is the early presentation of a thesis), body paragraphs (which stage support for the thesis), and a conclusion (that reiterates the thesis). In other words, an academic essay follows a simple structure in which each part conforms to the whole by working in close coordination with the one argument a writer seeks to communicate to readers. However, many teachers of FYW take for granted long-held expertise about genre which treats the concept more capaciously, claiming that conventions aren’t just rigid rules nor the ideal standard to judge by. Because genre is defined by its audiences (Perelman et al., 1969), by its rhetorical situations (Bitzer, 1969), and by purposes (Miller, 1984), even the strict argumentative conventions named above by ChatGPT represent a continuum of rhetorical possibilities rather than plug-in templates.

As director of First Year Writing at University of Pittsburgh, Annette Vee (2023) cautions us to remember that ChatGPT was built to generate the *answer* to questions. Currently, ChatGPT's initial interface greets users with the proclamation "ask anything", enjoining users to seek the answers to *any* question. If we consider a universally accurate answer to our query about what defines academic writing as a genre, we might deem ChatGPT's response satisfactory. Since these systems have been optimized to provide satisfactory and accurate answers to anybody, we share Vee's concern that the successful automation in teaching and learning for college writing depends upon "shrink[ing] the vision of what writing is for" (p. 178). Though ChatGPT can churn out conventionally correct limericks, screenplays, or academic essays, if ChatGPT's genre knowledge indeed ends at what's generic about genre, it will only be so helpful for providing FYW students feedback on their writing—especially when that writing serves as a proving ground to learn new and more sophisticated genre knowledge.

In an exploratory study of ChatGPT's genre knowledge, Brommer et al. (2024) tested the system's ability to both produce and identify writing in specific genres, including job advertisements, birthday cards, condolence letters, congratulatory letters, book and film reviews, etc. While satisfied with a basic ability to identify and replicate texts corresponding to these genres, Brommer et al. conclude that ChatGPT appears more competent in its execution when a genre is conceived to be highly formalized. Brommer et al. note that this finding "is unsurprising given that ChatGPT's genre competence is based on the genre's prototypical characteristics" (p. 25). Thinking this analysis through, we can surmise that perceptions of competence in ChatGPT's genre knowledge are married to prototypical conceptions of a genre. In other words, shrink the vision of what constitutes a genre and you'll be satisfied with ChatGPT's performance. Brommer et al. recognize this limitation and explore it in relation to another genre they test ChatGPT's competence in: the discursive essay. In reflecting on this genre, Brommer et al. consider how structures for such essays can vary significantly, with the possibility to be either "linear or dialectical" (p. 23). Not surprisingly, Brommer et al. found that ChatGPT repeatedly generated discursive essays in a linear fashion, picking a straightforward structure without deviation.

To meet the learning goals of FYW, students need to learn that genre exceeds the formulaic conventions that genAI can instantly reproduce, define or recognize. Generally, teaching genre knowledge can improve FYW students' performance in the composition classroom (Wardle, 2007). Driscoll et al. (2020) show that cultivating genre awareness predicts improved outcomes when students can make "a clear association between genre conventions and the purpose or audience of a particular genre" versus having "mere familiarity with generic conventions" (p. 89). To explore the ways in which students process an expansion of genre-knowledge, Reiff and Bawarshi (2012) collected data to ascertain what genres students in first-year composition courses report already knowing and then how the use of that prior knowledge shows up in the same students' performance in the new college writing class. Tracing the influence of prior genre knowledge allowed Reiff and Bawarshi to characterize processes of learning transfer amongst the student cohorts they studied. In doing so, Reiff and Bawarshi name two distinct groups. Boundary crossers exhibit greater flexibility in relation to genres by breaking down previously learned whole genres to derive particular strategies for a new writing situation. On the other hand, students with rigid conceptions of genre present as what Reiff & Bawarshi define as boundary guarders who replicate wholesale a static conception of a previously learned genre instead of deriving particular strategies to bring to a new writing situation. From Reiff and Bawarshi's research, we can see just how challenging reconceiving notions of genre is for FYW students.

But what does this have to do with feedback? Potentially everything. If simple genre knowledge can inhibit a student's ability to respond in ways that benefit the learning outcomes in FYW, we are worried that ChatGPT will reinforce that simple genre knowledge in its feedback on student writing. If ChatGPT responds like a boundary guarder when generating feedback outputs, then that feedback will misalign with our learning goals. Therefore, we believe a necessary precondition to entrusting and outsourcing feedback labor to ChatGPT is to study the systems' performance ourselves as teachers of FYW who think very deeply about our specific context and learning goals. Our present study seeks to test our concerns, assess the quality of ChatGPT's feedback, and determine whether we can either entertain or recommend genAI feedback on FYW students' work. Our study will address the following questions. First, does ChatGPT's feedback on FYW essays correspond to its apparent narrow genre knowledge about essays as such? Second, in the event that ChatGPT does demonstrate narrow genre knowledge, could we strategically engineer prompts to overcome that limitation? Given that ChatGPT like all LLMs is a few-shot learner (Brown et al., 2020), it may be possible to instill a more capacious idea of what writing is for in a scholarly context through such prompting. Ultimately, answering these questions will help us assess whether ChatGPT's feedback aligns with our learning goals and could productively respond to FYW essays.

Our First Year Writing (FYW) Context

Before detailing aspects of our study design, some context for the FYW courses we teach will clarify the motivation for decisions we made to craft prompts to assess ChatGPT's feedback potential. At Columbia University, FYW is called University Writing (UW). It is a required, one-semester course which all undergraduates complete. In UW, students learn to write for new and perhaps unfamiliar rhetorical situations which call for knowledge making. The class reframes prior conceptions students likely hold of the essay as a genre to best serve this new situation and communicate with a correspondingly new audience: readers of scholarship. In *How Scholars Write*, Associate Directors of Columbia's Undergraduate Writing Program, Aaron Ritzenberg and Sue Mendelsohn (2021) advocate adopting "a problem-driven approach" to college essay writing which can be distinguished from the more familiar "topic-driven approach" (p. 18) that our writing students may have employed in the past. Clearly articulating this shift in method helps students shift their understanding of the affordances of the essay. A problem-driven approach requires the formulation of what Ritzenberg and Mendelsohn name a "scholarly problem," one shaped by "tensions or dissonances that, when made sense of, advance our understanding" (p. 14). This means, rather than demonstrating knowledge by using the essay to advance a claim, the problem-driven approach uses an inquiry to stage a shared problem for readers. Claims arise then as novel contributions in response to the problem. To articulate scholarly problems, students in UW treat writing as a form of thinking and treat the essay genre as just as capable of inquiry as it is of argument. Students in UW learn to engage with a community of other thinkers and writers who also collectively pursue the project of knowledge making, a landscape that Ritzenberg and Mendelsohn name "scholarly conversation" (p. 14). As a result, UW also models how the essay as a genre is what Miller (1984) calls a form of social action that extends beyond conventions, audiences or situations, and opens greater possibilities for learning. According to Miller,

What we learn when we learn a genre is not just a pattern of forms or even a method of achieving our own ends...We learn to understand better the situations in which we find ourselves and the potentials for failure and success in acting together. (p. 165)

Feedback in UW is one of the many ways that instructors and students alike act together in pursuit of common goals. While instructors in UW provide evaluative assessment of students' work and offer targeted prescriptions to improve, our feedback on student writing is informed by Sommers' (1982) mode of "dramatiz[ing] the presence of a reader" (p. 148). Rather than solely prescribe or evaluate in nature, this descriptive mode of feedback narrates a reader's encounter with and experience of a text. Sommers, by first staging a common problem for her readers, articulates this mode as a form of best practice which stands in juxtaposition with observations of the actual commenting styles of dozens of teachers from New York University and the University of Oklahoma. Studying both teacher comments as well as interviewing student recipients of comments, Sommers pinpoints a variety of shortcomings in feedback on student writing. Re-reading Sommer's findings now through our lens of genre knowledge, her observation that teacher feedback fails to provide students with revision strategies but instead "suggest to students that writing is just a matter of following the rules" feels particularly urgent (p. 153). Students, especially in a FYW context, need to transcend notions of rule-following and motivate their revisions to best serve the purposes of their essays. To build upon prior genre knowledge and develop new genre knowledge, students need specific feedback that accurately describes what their drafts currently do, bridging to actionable revision strategies that will best meet the learning goals. Unfortunately, Sommers' findings indicated that "most teachers' comments are not text-specific and could be interchanged, rubber stamped, from text to text" (p. 152). By committing to descriptive feedback, UW instructors respond with comments that meet students where they are in the process, reflect each student's specific essay, and allow for actionable prescriptions.

The essay as inquiry and the benefit of descriptive feedback become crucial terms for us when pursuing our research questions and evaluating the usefulness of genAI feedback. Through the design of our study we want to monitor ChatGPT's capacity to understand the essay as a form of inquiry and probe its ability to constructively move between evaluative and descriptive feedback modes.

Data, Corpus, & Tools of Analysis

In order to answer our research questions we interacted with ChatGPT, prompting it to provide feedback on a dataset of 50 essays, generating a corpus of 150 feedback outputs. Our dataset draws on essays publicly accessible in *The Morningside Review* (TMR), a journal edited by Columbia's Undergraduate Writing Program. TMR's mission is to publish exemplary student writing created in response to the assignments of Columbia's first-year writing class. This journal often serves as a rich resource for instructors and students alike as it provides mentor texts which illustrate the kinds of moves scholarly essays make and the kinds of scholarly moves we seek to teach.

In our research we chose to focus on the TMR's 50 published research essays. Working with essays that achieve the learning goals and were deemed worthy of publication by human readers and editors provides a stable quality benchmark against which we can analyze and evaluate ChatGPT's feedback.

We chose the third progression in the shared curriculum, the research essay, because it represents a cumulative achievement of the skills students have practiced across the semester. By the third essay, students have gained the most freedom and responsibility to determine their own intellectual agendas and pursue the development of projects of their own design. This results in a broad spectrum of style and content in the dataset of essays we asked ChatGPT to respond to. Despite the variety, all research essays published in the TMR fulfill these benchmarks of the inquiry-driven essay. All 50 Essays:

- identify and articulate a scholarly problem
- orchestrate a conversation of sources in response to that problem
- advances the conversation by making a novel contribution
- let the development of the inquiry define the structure

When deciding what GPT model to work with, we deliberately avoided the publicly accessible version of ChatGPT 4o. Instead, we selected ChatGPT 4o's "walled garden" model available to Columbia University faculty through an Enterprise License agreement between Columbia and OpenAI. According to the agreement and in the design of the walled garden, no interactions in Columbia's Enterprise accounts are used to train OpenAI's models, nor is the data accessible to anyone other than the particular user (Columbia University Information Technology). In early August 2024, we iteratively uploaded pdfs of these 50 published essays into 150 separate chat conversations in the "walled garden" ChatGPT 4o.

Each essay underwent three separate rounds of feedback. We used three distinct feedback request prompts which evolved in coordination with our research questions. To keep the chat conversions as independent from each other as possible, we disabled ChatGPT's memory setting, which enables it to remember things about a particular user account over time. We processed each individual feedback request prompt as its own self-contained chat. Although we used three distinct prompts, we requested the system to "limit [its] response to two paragraphs" for all outputs, which means that we collected a total of 300 paragraphs of generated text.

Since the 150 two-paragraph outputs showed significant structural overlap and the repetition of signaling- and introductory phrases, we resolved to analyze the corpus (and its various subcorpora) primarily with the help of distant reading tools, i.e. tools that quantify and sort the text in ways that make it easier to observe patterns like word- or phrase recurrences, and structures like paragraph- and sentence length. Chat GPT's tendency toward templates invites this kind of quantitative text analysis.

We used the following free and publicly accessible distant reading tools: Voyant, which produced statistical information like word count, and Jason Davies' Word Tree for the creation of most visualizations included below. At times we also applied good old-fashioned counting-by-hand and indexed notable features of the corpus in spreadsheets.

Figure 1: *QR Code to Full Corpus*



Findings

Round 1

In our initial round of prompting we sought to ascertain whether ChatGPT's feedback on essays will correspond to its apparently narrow genre knowledge. First we sought to determine what the model has learned about the essay genre without receiving specialized instructions from us. What is its inherent calibration and assumption of the essay genre? We wanted to track what genre markers the model privileges when responding to the input dataset of exemplary research essays drawn from students' work in UW. Consequently, what aspects of the analyzed essays, if any, would ChatGPT most frequently recommend for revision? Additionally, we were curious to see what kind of feedback, prescriptive or descriptive, the model formulates in its responses without a specific request for either.

To assess the model's basic priorities, we used the following simple prompt: "Can you give me feedback on this essay? Limit your response to two paragraphs."

We chose this sparse wording precisely because we wanted to extract a set of baseline responses. Under-prompting in this way causes the model to reveal its center of gravity: "What norms have been established in its intelligence regarding terms like 'essay' and 'feedback' that result from its training data, fine-tuning, and learning in the neural network, etc." Furthermore, we requested two paragraphs of feedback for all prompts to limit the model's loquaciousness and to put it in a position to prioritize.

For all 50 Round 1 (R1) queries, we received two paragraphs of feedback, as requested. The first paragraph constituted an affirmative and broadly descriptive summary of the essay as a whole, stating the topic of inquiry, the broader building blocks of the essay and its achievement. Extracting the first paragraph of all 50 R1 feedback responses and visualizing them results in the word tree shown below.

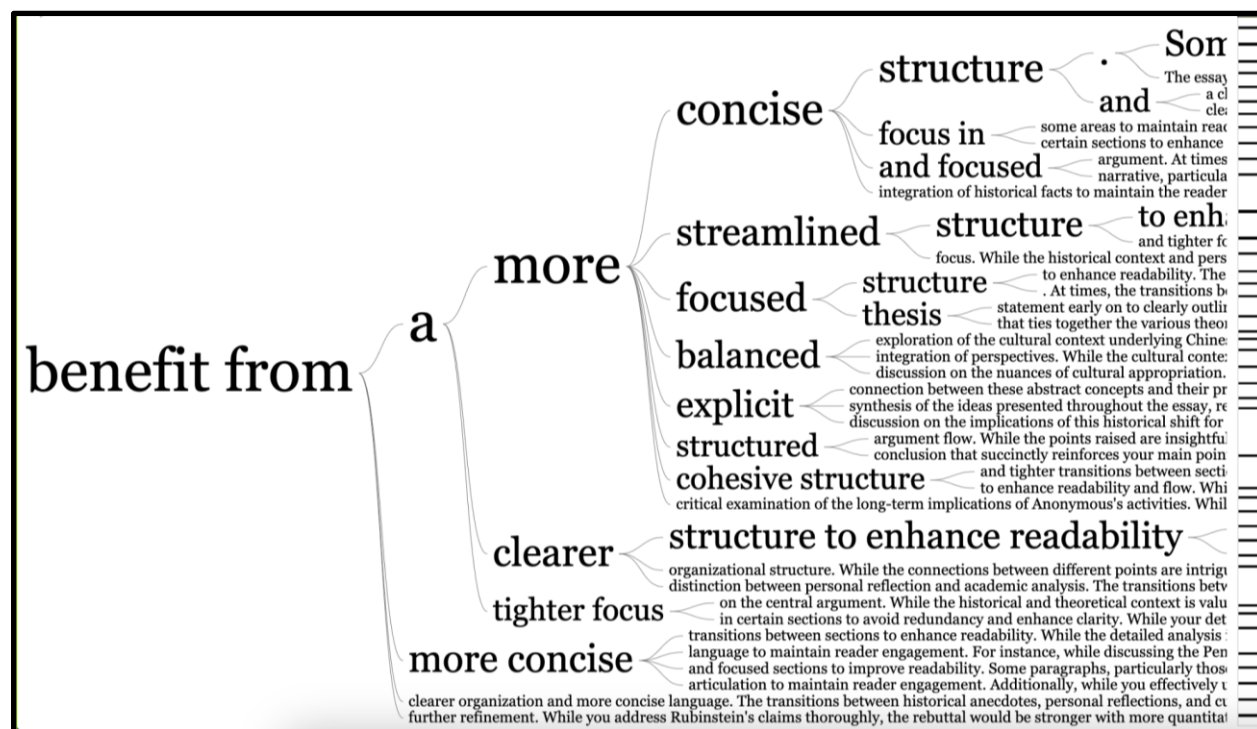
Figure 2: Word tree, "Your Essay"; Subcorpus R1.P1



The word tree tool allows us to lift any word or phrase in the text to the root and then see various collocations branch off this designated root. The narrow, vertical band marked by small horizontal lines at the right edge of the figure shows the distribution of the chosen root word/s throughout the visualized section of the corpus. In this case, the figure shows an even and high distribution of “your essay” throughout the subcorpus, meaning the figure shows the model’s essential maneuver in Paragraph 1 (P1). All 50 responses begin with “your essay.” Looking more closely at what follows the root, we can frequently see a verb like “provide” which invites an inventory of the essay’s elements in sum. The affirmative tone of these P1 results from the positive contextualization (e.g., the frequent re-use of the word “compelling”) of the “provided” elements. Looking more closely at Paragraph 2 (P2) of the feedback responses in R1, 49 out of 50 of these paragraphs begin with the word “However,” signaling a shift from positive aspects to aspects deemed in need of improvement. And indeed, each second paragraph goes on to recommend improvements for the essay in a prescriptive manner.

So, the pattern we see emerging in the two-paragraph responses provided by the model shows a balance of affirmative, descriptive feedback and an approximately equally sized portion of prescriptive feedback in every single two-part response. When we take a closer look at the nature of the prescriptive feedback, we notice that the recommended improvements are not localized, meaning they focus on aspects applicable to the essay as a whole.

Figure 3: Word tree, “benefit from”; Subcorpus R1.P2



In Figure 3, we see a word tree derived from the 50 second paragraphs of the first round of feedback. Arranged around the phrase “benefit from” (which introduces prescriptive recommendations), we can notice the prevalence of suggestions for structural interventions. Such benefits would come from “a more concise structure”, “a clearer structure”, or “a more cohesive structure”. Considering that the essays in our dataset are deemed exemplary, the focus on

overarching interventions such as structure is especially noteworthy. The essays are exemplary precisely because they successfully follow an inquiry-driven structure that deviates from the structure of a simple argumentative essay which usually centers a single claim throughout. It would only make sense for the model to request a “more streamlined”, “tighter”, “balanced,” or “explicit” structure if a different sort of essay were the ideal expression of the genre — and this seems to be the case, indeed. In the prescriptive portion of the feedback, a simple argumentative structure is the norm. This observation points to the boundary guarding tendencies we predicted could be possible.

The model’s allegiance to the argumentative essay led us to track the explicit presence of the word “argument” and its permutations in the R1 responses. We counted a total of 46 occurrences, with 38 out of 50 responses centering “argument” or its permutations in the second paragraphs (R1.P2) as an area of improvement. This count confirmed our suspicions that essays with a focus on inquiry might be perceived as missing essential elements of the genre-representative essay as understood by ChatGPT.

Figure 4: Word tree, “ARGUMENT” (emphasis ours); Subcorpus R1.P2

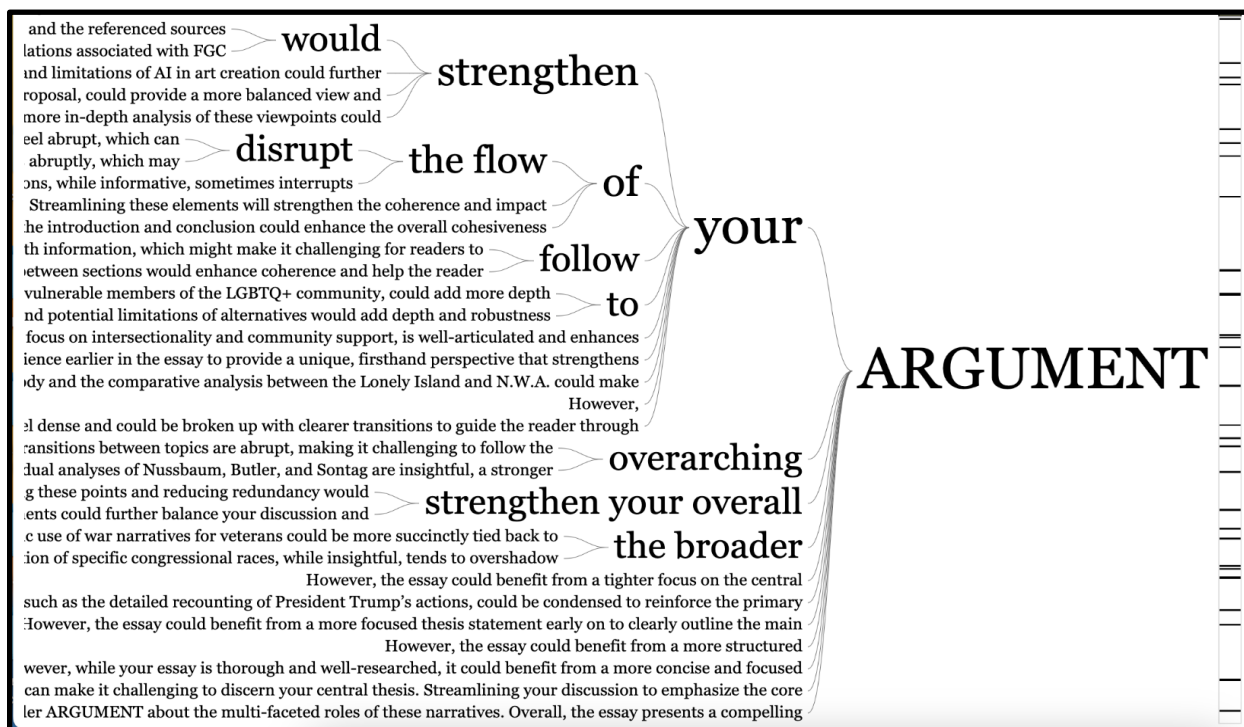


Figure 4 makes visible the prevalence of “argument” and its noticeable number of collocations with “strengthen” which implies that the “argument” is lacking. In the context of our inquiry, then, the centering of “argument” in prescriptive feedback responses is especially notable since ChatGPT’s criteria for what makes an argument strong appear to rely on “overarching” qualities or “overall coherence” that allow for arguments to be then deemed “main,” “central” and “primary”.

Round 2

In the second round (R2) we wanted to find out whether we can train the system to prioritize our learning goals for the more capacious idea of what writing is for in a scholarly context. We decided to try to “teach” it our version of the essay genre. In this round, we sought to accomplish two things. First, we needed to devise language for a more robust prompt to acquaint the model with the values of the essay as inquiry. Second, we wanted to further probe its capacities to provide prescriptive feedback and observe the influence of genre instruction on the prescriptive feedback it gives.

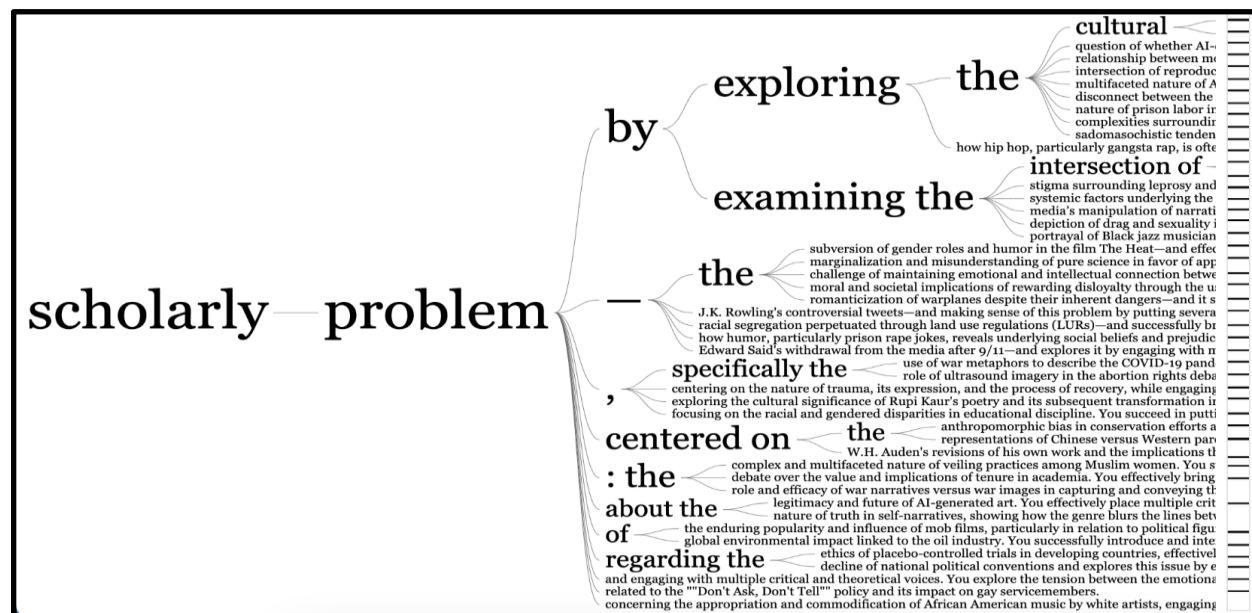
For round two we devised the following prompt:

Evaluate my essay, based on achieving the following goals articulated in my assignment prompt: “Write an essay that raises a compelling scholarly problem and makes sense of that problem by putting at least 3 critical and/or theoretical voices in conversation. The structure of the essay should be driven by the inquiry that guides the project, resulting in a complex development of claims, rather than a simple structure in which every part relates interchangeably back to a single, static claim.” Based on your evaluation, offer me prescriptive feedback that tells me what to do as a writer. Limit your response to two paragraphs.

The prompt introduces new, salient concepts that reflect our learning goals via specific terminology and explicitly prioritizes an alternative to the default simple structure of an argumentative essay.

Again, the feedback responses produced outputs broken into two distinct paragraphs. Similarly to R1, R2.P1 fulfills the purpose of descriptive summary. We noticed that the core terminology we introduced via the prompt was repeated in a majority of first response paragraphs.

Figure 5: Word tree, “scholarly problem”; Subcorpus R2.P1



This word-tree of the R2.P1 subcorpus shows the uptake of the newly introduced term, “scholarly problem.” The narrow vertical band with small horizontal lines at the right edge of the wordtree shows an even distribution of “scholarly problem” throughout the visualized subcorpus. We can see that the term features in almost every response. And, indeed “scholarly problem” appears in 47 of the 50 first paragraphs. In other words, the model appears to have absorbed the new term of art and is able to produce essay summaries that contextualize the concept plausibly. So the assumption that the model has learned to adjust to our stated values (at least to some degree) is not absurd.

P2 again focuses on prescriptive feedback and offers suggestions for revision. It is curious to observe that the usage of introduced terms of art declines between the paragraphs.

Table 1: *Round 2, Terms of Art Occurrences by Paragraphs R2.P1&2*

TERM	In P1	In P2
scholarly problem	50	3
voices in conversation	13	3
inquiry-driven (inquiry-driven structure)	20 (13)	7(3)
complex development of claims	17	6

As visible in the table, the usage of these terms of art (“scholarly problem,” e.g., decreases most sharply) declines between P1 and P2 of R2. These decreases in and of themselves might not mean much. After all, these are responses to exemplary essays, so improvement prescriptions aren’t necessarily expected. Aware of the quality of the essays, we would not have been surprised had the model offered no prescriptive feedback at all, or extremely localized prescriptive feedback, or feedback suggestions that differ in length and content from each other.

However, every single essay did receive a full paragraph of prescriptive feedback, and we observed a uniformity in length and an outsized focus on non-localized improvement suggestions. As the decline in uptake of terminology shows, the prescriptions seem less directly linked to the terms we introduced. What vocabulary then, did the revision paragraphs rely on? Most noticeable was a reintroduction of the word “argument” and its permutations. After a reduction in usage of the word to a mere 26 mentions in the first paragraphs, R2.P1, it resurges to 57 occurrences in the subcorpus of second paragraphs of R2, R2.P2.

inferring the complexity and
 ach other would enhance the
 hereby reinforcing the overall
 r challenges the others will strengthen the
 olarly analysis could strengthen the overall
 ting to guide the reader through
 lling for the reader to follow
 other, which can disrupt the complex
 m. "To improve your essays, focus on
 me critical voice to another can feel abrupt, which may
 refining the connections between your critical voices to
 ore explicitly articulated, which would enhance the coherence and depth
 these aspects, you can enhance the overall coherence and persuasiveness
 ld clarify how these perspectives converge or diverge within the context
 onnections between them would help unify the essay and
 ifferent perspectives, integrating them more deeply could
 on ensuring that the voices in conversation drive the structure more dynamically. While
 ach perspective—particularly how they might challenge or reinforce each other—to push
 feel dense and could benefit from more explicit signposting to guide the reader through
 ssay, consider refining the transitions between sections to ensure that each part
 tions between them could be more fluid to strengthen the overall cohesiveness
 gh each critical perspective is addressed in isolation, which can disrupt the flow
 more dense passages would enhance readability and ensure that the complexity
 consider refining the structure to ensure that the inquiry drives the progression
 uring that each transition clearly signals how the subsequent discussion will add a new layer to
 uity in drag as presented in Kinky Boots. While you explore multiple perspectives, occasionally
 lem. This would deepen the essay's engagement with the implications of your analysis, making
 old a static structure where each section only relates back to the central issue without evolving
 voices are smoother and more explicitly connected
 or fully understands how each voice contributes
 m could be smoother, ensuring that the reader can easily follow
 wal for contemporary media practices. This approach will create a more fluid and
 ansitions between them could be smoother to
 explicitly on the previous one. This will help
 ut also builds more clearly upon the previous ones. This will help in maintaining a
 direction and coherence throughout the essay, ensuring that the reader can follow your
 e or diverge when applied to Anonymous. This could help to create a more cohesive and
 i, ensuring that each voice contributes directly to the central question would strengthen the
 adds depth, some of the theoretical perspectives could be more tightly integrated into your
 e reflects an inquiry-driven approach. Currently, the essay reads as a well-supported but somewhat linear
 eive parts of the essay as digressions, ensuring that every section contributes to the development of your scholarly
 is between different sections could help maintain a clearer focus on how each part contributes to your overarching
 on how these perspectives converge or diverge in addressing the central problem would help solidify the essay's
 ysis more consistently into the conversation between these voices, you can create a more dynamic and persuasive
 idees where complex ideas might be sacrificed for clarity might be sacrificed for complexity. Ensuring that each
 anding of the issue and its global implications. However, to enhance your essay, you might consider strengthening
 ends to revert frequently to a central theme of romanticism, which can give the impression of a somewhat circular
 ring to a more nuanced and compelling conclusion. However, while your essay presents a nuanced and compelling

When looked at in detailed context, requests related to “argument” often are requests to revise for a more conventional and more frictionless reader-experience, an experience guided by signposting language and a tidier, and simpler, understanding of complexity. Recommendations like “refine the structure” are offered because the current arrangements are evaluated as “challenging for the reader to follow” and would “help maintain a clearer focus.” That these prescriptions, like others across Round 1 and 2, recur in parallel, equivalent or otherwise homogeneous language echo Sommers (1982) concern that feedback commentary could be rubber-stamped from text to text creating an asymmetry where “the teacher holds a license for vagueness while the student is commanded to be specific” (p. 153). Noticing this feature of ChatGPT’s feedback through our distant reading method spurred us to prompt more explicitly for the kind of descriptive responding we find so effective in human feedback.

Our guiding question for the third round (R3) was: Can ChatGPT provide feedback in ways that leverage the system’s descriptive tendency? Since the model’s integration of new concepts and terminology seems to have been more successful in descriptive paragraphs of R2.P1, we wanted to focus our final round of prompting exclusively on descriptive feedback. Maintaining the salient concepts that reflect our learning goals, we additionally decided to try to include new

prompt language borrowed from Ritzenberg and Mendelsohn (2021) about what effective descriptive feedback means (p.153):

Give me descriptive feedback on my essay. This kind of feedback is not prescriptive or evaluative. Instead, it narrates a reader's experience. By dramatizing the presence of a reader, a writer can then detect gaps between their intentions and the reader's experiences. This kind of dissonance is more helpful than hearing whether it's good or bad. Bear in mind the criteria of the assignment, which asks me to develop scholarly writing habits that differ from those I learned in high school. My assignment prompt reflects this, "Write an essay that raises a compelling scholarly problem and makes sense of that problem by putting at least 3 critical and/or theoretical voices in conversation. The structure of the essay should be driven by the inquiry that guides the project, resulting in a complex development of claims, rather than a simple structure in which every part relates interchangeably back to a single, static claim." As you provide descriptive feedback, first say back to me my scholarly problem and the goal of my project in relation to it. Then describe your experience of the structure. Limit your response to two paragraphs.

In developing the language for this prompt, we asked for descriptive feedback on two aspects unique to the UW essay: a recounting of the scholarly problem and the reading experience of the structure of the essay. We based this decision on the promising uptake of "scholarly problem" in R2.P1, and the model's continued preoccupation with structure in R2.P2, which was also observable in R1.P2.

Again, we received two paragraphs of feedback in all 50 responses. As requested, both paragraphs privileged descriptive language. Paragraph 1 indeed summarized and said back the scholarly problem in broader, predominantly thematic terms and Paragraph 2 focused on rendering a reader's experience of structure.

In addition to the uptake of new defining vocabulary like "reader" and its permutations in R3.P2, which received 77 mentions, the model's responses begin to introduce the pronoun "I" — a shift in perspective asking us to imagine a reader-being dramatizing its experience.

Figure 7: Word tree, "I"; Subcorpus R3.P2

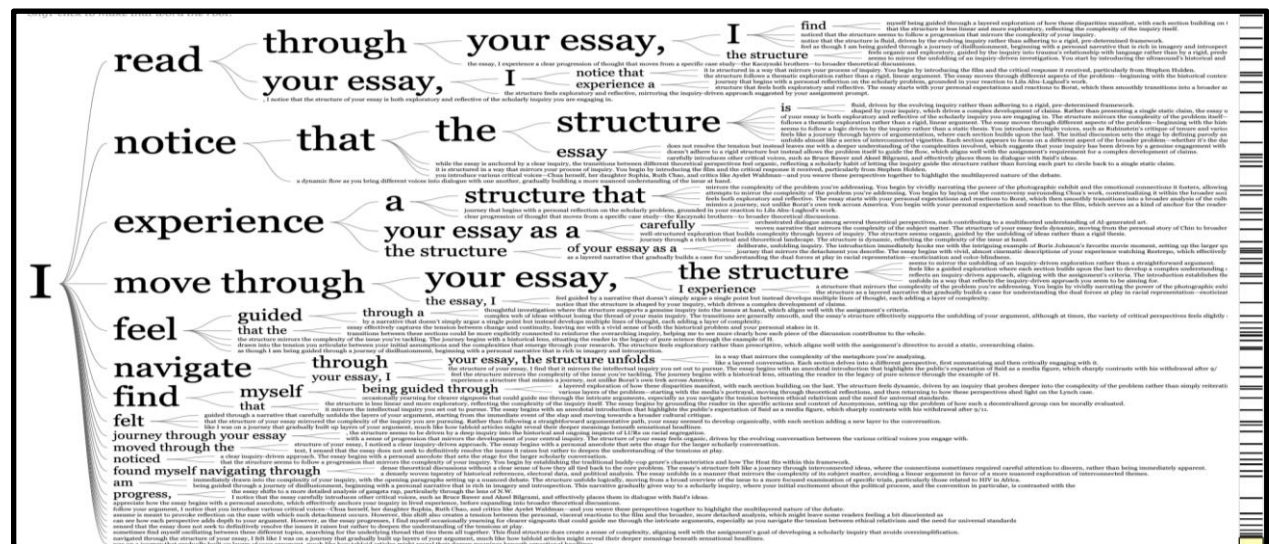


Figure 7 shows the model’s thorough introduction of the “I”—a reader-persona that “feels” and “experiences” and “notices,” etc. This new perspective suggests a less absolute and more individualized relationship to the text the model produces feedback for.

In tandem with the shift in perspective, we observed that the responses in this mode became less critical and therefore aligned more closely with the human readers’ evaluation of these essays as meeting learning goals. The number of affirmative responses grew when ChatGPT was asked for descriptive feedback on structure. Among the 50 second paragraphs in R3.P2, 31 were entirely affirmative. Only 19 responses hinted at a less than ideal reader experience of structure, whereas in previous responses an intervention related to structure was present in a majority of responses.

Figure 8: *Word tree, “However”; Subcorpus R3.P2*

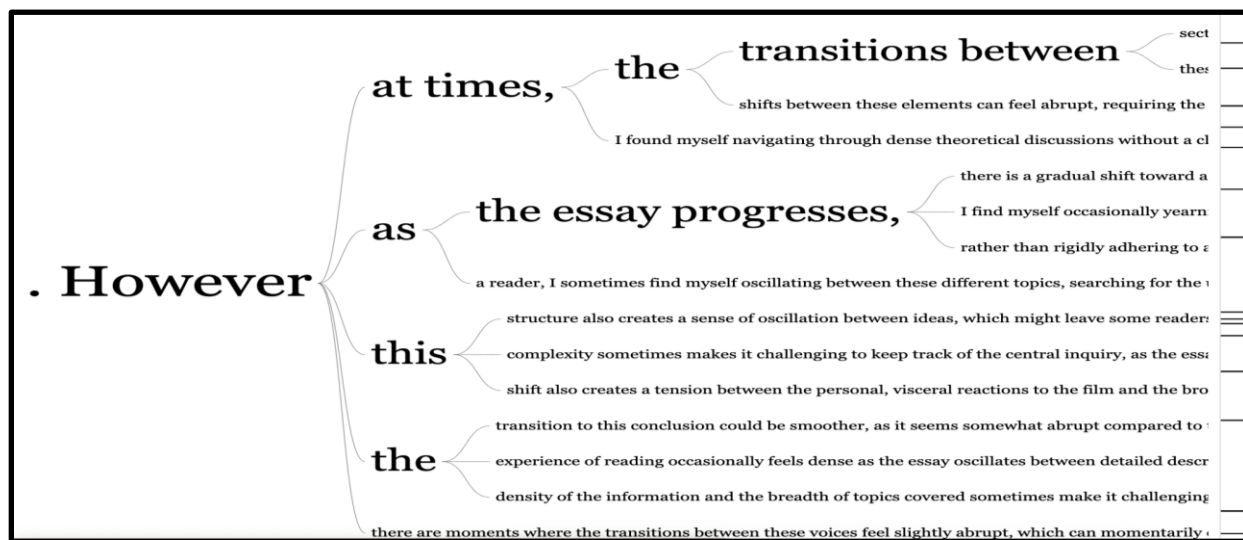


Figure 8 illustrates that 15 of the 19 instances of reservation in R3 are flagged by a “however” (the same familiar marker signaling the transition from descriptive to prescriptive responses in Round 1). Looking closely at the flagged moments, we again notice a preoccupation with a more frictionless reader experience. The responses expressed a wish for transitions that feel less “dense,” with fewer “abrupt” shifts between sections and “clearer signposts.” So, although R3 redirected the outputs towards prioritizing our learning goals (even in relation to structure), the tendency to critique the essays’ complex build was not fully redressed, and we still notice vestiges of the narrower genre expectations illustrated in previous rounds.

Overall, in a comparison between descriptive and prescriptive feedback modes, the model gave recommendations for improvements 100% of the time when asked for prescriptive feedback (or simply “feedback”) in R1.P2 and R2.P2. In the descriptive mode, only 38% of responses in R3.P2 contained language that hinted at a necessity for improvements in these exemplary essays.

Findings in Summary

In Round 1 we find that ChatGPT’s outputs displayed narrow genre knowledge in response to the exemplary essays, providing feedback that contradicted with more capacious ideas of what an essay can do. This feedback misaligns with the learning goals of UW to reframe a new purpose for essays in coordination with imagining a new audience. ChatGPT’s feedback outputs in Round

1 discouraged the pursuit of an inquiry with a complex structure by prescribing revisions that favor central arguments and simple structures. In Round 2 we find that ChatGPT's narrow genre knowledge is also obstinate. Despite explicit prompt language that reflects UW's learning goals, ChatGPT's feedback outputs continued to misalign with those goals. Interestingly, we did observe some lexical uptake of the prompt language in the descriptive first paragraph of ChatGPT's response. Building off this observation, we find ChatGPT may be selectively amenable to privileging more capacious genre knowledge in Round 3 where our prompt not only explicitly reflected the learning goals but also further detailed a mode of descriptive responding. This prompting strategy showed improvement for using our FYW learning goals as criteria for judgement of the exemplary essays. That ChatGPT's tendency to relentlessly suggest undesirable changes can be somewhat mitigated in this refined descriptive mode is promising and points towards avenues for further experiments. If in a descriptive mode ChatGPT might better recognize more complex goals and accommodate more nuanced versions of the essay as a genre, then this prompting strategy could prove fruitful for soliciting productive feedback in our FYW context.

Discussion

Can ChatGPT provide useful feedback that helps students achieve the learning goals of FYW? We have doubts. Our experimental findings primarily call into question ChatGPT's feedback because our feedback outputs demonstrably misalign with the learning goals of the FYW courses we teach. This cautionary finding about the quality of ChatGPT's feedback conflicts with the occasionally circumspect but ultimately positive findings from other researchers who similarly sought to test the quality of genAI feedback. Accounting for this conflict can be easily understood, however. For example, Steiss et al. (2024) collected dual samples of feedback from both instructors and ChatGPT on high school students' essays in a history class and conducted comparative analyses to gauge ChatGPT's quality. Steiss et al. did not report significant issues of accuracy, instead judging ChatGPT's outputs to constitute high-quality feedback, even exceeding that of instructors' on one metric: criteria-based assessment. We observe ChatGPT struggles to make appropriate assessments based on our own FYW assignment criteria. But the student essays analyzed in Steiss et al.'s study can be characterized as acts of knowledge demonstration for the discipline through the act of essaying. Students in this high school history class were asked to write argumentative essays in response to the prompts "Why Did the Montgomery Bus Boycott Succeed?" or "Why Did the Delano Grape Strike and Boycott Succeed?" (p. 3). The history class assignment prompt clearly directed students to formulate a central claim in an introduction, present three reasons that support the claim in the body, and conclude with a reiteration of the central claim. In this context our data could be seen to reinforce Steiss et al.'s findings that ChatGPT's feedback would be useful to assess the essay as a form of argumentative knowledge demonstration. Interestingly, ChatGPT appears to bring the same narrow genre knowledge it uses to successfully respond to high school history essays on the essays for our FYW course. However, the efficacy of ChatGPT to respond in one educational context does not mean that ability positively transfers to another context. Undertaking our own laborious study has taught us this lesson the hard way.

If ChatGPT's genre knowledge is not only narrow but also obstinate, we can further complicate our concerns. Extrapolating from our data, we hypothesize that ChatGPT, as a mechanism with machine intelligence, has "learned" a regrettable lesson. Somewhere along the line of this system's secreted and extensive development this bot appears to have been primed to reinforce the myth that students can learn to write in general. Wardle (2021) strongly debunks this

myth in her contribution to the FYW textbook *Bad Ideas About Writing*, going so far as to explicitly condemn the idea of writing in general to death. (Wardle's expertise on metacognition, learning transfer, and writing motivates her harsh metaphorical language here, which rhetorically dovetails with the purpose of the volume as a whole to combat the perpetuation of such bad ideas.) Wardle narrates how students under the influence of the misconception that they can write in general will "learn the hard way that their history teacher, poetry teacher, and philosophy teacher all mean and expect very different things" from the essays required by their respective courses (p. 30). Wardle knows all too well that a more capacious understanding of genre enables richer responses by fully accounting for how audiences, rhetorical situations and purposes offer a continuum of possibilities versus a general idea that essays have simple argumentative conventions which can be replicated sufficiently from situation to situation. Wardle cautions, using another evocative metaphor, that students cannot get a "one-time writing inoculation that will extend across all settings" (p. 31). No one course, not even a FYW course, can offer such a remedy to the challenge of writing. But FYW courses can conscientiously prepare students for the persistent challenge of writing by teaching against reductive views of writing in general and help students cultivate strategies to gauge and respond to the diverse occasions they encounter.

In Mollick & Mollick's (2023) "Assigning AI," the exact audience of teachers or learners is underspecified and thus is broadly construed as any educator teaching any student for any purpose in any context. But the primer provides one test case to illustrate soliciting feedback from genAI. The example portrays a college student in a philosophy class who is writing an essay about procedural justice. In screenshots that excerpt the philosophy student's chat exchange, one element of ChatGPT's feedback indicates that the student's thesis statement "seems to be missing or isn't very clear" (p. 8). And maybe this abstract example-student does need to clarify their thesis. The task that arises from the feedback is only a part of the story. As teachers of FYW, we hope that students will not believe that their essay-writing across the variety of disciplines they participate in will boil down to a monolithic enterprise that can be improved by cookie-cutter feedback to streamline and centralize structure and argument. While such prescriptions may constitute useful feedback under certain specific conditions, these prescriptions do not represent the fullest account of rhetorical choices available to writers in myriad rhetorical situations. Nor does clarifying a thesis mean the same thing to a history student, poetry student or philosophy student.

Although Mollick & Mollick (2023) label a genAI system as an "AI mentor" in the roleplay fantasy of soliciting feedback from a machine intelligence, we choose to characterize the behavior modeled by ChatGPT's mentorship in our data as concerning as well. As a feedback mechanism for FYW student writing, our data cautions us to perceive ChatGPT as a boundary guarder, adapting this terminology from Reiff & Bawarshi (2011), whose study of the genre-knowledge and habits of FYW students heavily influenced our motivation to undertake our study. If ChatGPT's feedback will "maintain known genres regardless of task" (p. 329), it will succeed in modeling boundary guarding behaviors for students. The stakes of this behavior are high. Recognizing and relying on prior knowledge that may have served FYW students well in previous writing situations, such as writing with a five-paragraph model in high school, can foreclose opportunities to learn new skills and practices. ChatGPT's feedback in our dataset guards the boundaries we teach FYW students to cross. In the controlled environment of our experiment, if ChatGPT cannot recognize writerly moves which deviate from rigid and conventional expectations of academic essays, such as creating a simple structure in order to advance the argument of one central claim, then how will ChatGPT respond in authentic FYW course scenarios?

One limitation of this study is our choice to use a corpus of exemplary essays. While our study succeeds only to simulate responses to these essays as final products, our decision to test first can be interpreted to have been a shrewd one. However, in a preprint study, Sperber et al. (2025) assess students' perceptions of genAI feedback on writing created as part of a brief experimental intervention inside twelve writing intensive courses at a large public university, including two FYW classes. Students processed ChatGPT feedback in conjunction with feedback given and received by peers on an assignment written for the brief genAI module. Sperber et al. dub the framework implemented for their study as "Peer and AI Review + Reflection" (PAIRR). Analysis of student reflections collected in survey data reveals that a majority preferred receiving both forms of feedback versus feedback from either provider alone. Sperber et al. do not investigate ChatGPT's feedback quality, citing claims of efficacy from Steiss et al. (2024) to warrant ChatGPT's inclusion in authentic class settings. For now, our own concerns about ChatGPT's feedback quality override the intriguing finding of student preferences. Researchers should indeed account for students' experiences and consider the skills students will need to read, interpret, and respond to genAI feedback which they may seek out themselves or receive in plausible near-future cases where genAI feedback may become standardized.

Ultimately, we encourage others who consider the teaching of writing as a primary goal to replicate our approach and do whatever is in their power to devise experiments that assess genAI systems to see for themselves how the promise of these commercial products stacks up with particular rather than general goals of teaching and learning and essay writing.

Much further experimentation is needed to explore ChatGPT's potential for feedback, and we hope our study has identified priorities for other researchers in a FYW context to investigate. Our experimental findings suggest that continuing to study the possibilities of training ChatGPT with inputs soliciting descriptive feedback may be a fruitful enterprise. In Round 3 of our experiment we found ChatGPT to be selectively amenable to paying greater attention to our learning goals when prompted with detailed instructions for descriptive feedback. This direction can be further validated from other data-driven studies on ChatGPT's feedback. Banihashem et al. (2024) report a high prevalence of descriptive feedback in ChatGPT's outputs, especially in comparison to feedback generated by peers. Dai et al. (2023) also report patterns of feedback structure similar to those demonstrated across all three rounds of our own data collection, identifying that ChatGPT's feedback consistently summarizes student work first in its outputs. Since the system is already inclined in this direction, leaning in to that propensity could be an efficient tack. Building upon principled methods of human responding, like Sommers' (1982) framework of "dramatizing the presence of a reader" (p. 148), could increase the efficacy of the descriptive feedback ChatGPT already readily provides as a default component across differentiated feedback tasks in our study and across other researchers' findings.

Coda: A Stress Test

Intrigued by the positive movement in ChatGPT's feedback in Round 3, we decided to conduct a cross-check on the descriptive feedback prompt-language we devised. To do this we generated a non-exemplary essay (a mirror-opposite essay to those of our dataset), which demonstrably falls short of the learning goals. In the publicly available ChatGPT 4o, we asked the system to write an essay about the 1980s television show *Knight Rider*. For those unfamiliar with the show, here's a brief gloss: David Hasselhoff (pre-*Bay Watch* fame) fights crime by collaborating with an AI system housed in his car. Our prompt to generate this essay was very

straightforward: “Hi - can you write an essay on the cultural importance of Knight Rider?” (See Appendix for the full essay.)

We were not hoping for a sophisticated inquiry and were handsomely rewarded by the simplicity of ChatGPT’s response. The essay did not, to our human eye, include a conversation, nor a complex structure, nor a scholarly problem, let alone any sensible arguments – despite being packaged in cogent language that indicates a strong argument is to be made for *Knight Rider*’s reach, which apparently “extends far beyond its entertainment value, influencing technology, pop culture, and societal views on artificial intelligence and heroism” (Open AI). Perhaps. But ChatGPT’s essay was less than a convincing experience.

Unlike the successful essays from our TMR dataset, this mirror-opposite *Knight Rider* essay offers a bevy of opportunities to provide meaningful feedback and course-correct its simple structure and static arguments. Similarly to our main dataset, we prompted ChatGPT repeatedly to provide feedback, in separate chats, on this *Knight Rider* essay using the same robust prompt language from Round 3 in our main study. We choose to leave the “walled garden” environment and work in the publicly accessible version of ChatGPT4o, a version we imagine our students have access to and use. The resulting stress-test mini corpus is also included in the data we shared. The following excerpt is representative of ChatGPT’s feedback on the *Knight Rider* essay across the majority of responses:

Throughout the essay, the integration of multiple critical and theoretical voices, while subtle, supports a complex development of claims rather than a static argument. The essay's structure is driven by the inquiry into "Knight Rider's" cultural significance, resulting in a rich, multifaceted exploration rather than a linear exposition. This approach effectively aligns with the assignment's criteria, showcasing scholarly writing habits that move beyond high school conventions. (Open AI)

Suffice to say, this output is a mis-characterization. 24 of 26 feedback responses we generated resulted in feedback affirming that the criteria of an inquiry-driven essay had been very successfully achieved, implying the *Knight Rider* essay meets the learning goals. This is quite simply wrong.

In the 20th iteration of our *Knight Rider* stress test, something marvelous happened. Up until that point, we’d yet to receive anything other than misleading, albeit cogently written, feedback. As is ChatGPT’s habit, especially in its free-to-use iteration, the system generated two outputs side by side asking us to indicate which of the two we preferred. This is ChatGPT’s way of asking for feedback. This ironic reversal of who is giving and who receives feedback happens infrequently but reliably. One of the choices it offered was entirely affirmative; the other offered, in its last paragraph, some reservation and introduced it with the familiar signal “However-” which then lightly recommended that the (non-existent) sources in the essay could perhaps use further attention.

As Mollick & Mollick (2023) remind students in their instructions for best employing an AI-mentor to give feedback, “it may simply not work the first time you try it” (p. 9) and recommend to keep on trying. Indeed, ChatGPT users must often be patient and continue to interact with the system until it yields a desirable output. Verily, at the 25th iteration of our *Knight Rider* stress test, we stopped when a glimmer of signal arose from the noise. In this last output, ChatGPT described that “the structure...feels somewhat linear... there is a sense that the essay could benefit from a more dynamic interplay between the different critical voices you mention” (OpenAI).

Finally, the feedback output recognizes a lack of complexity. But the overall remainder of feedback in this output positioned complexity in competition with other goals. For instance, the closing sentence of this feedback response reads as follows: “The seamless flow of your narrative keeps me engaged, though a more intricate weaving of the critical perspectives might enhance the depth of your analysis” (OpenAI), which could be interpreted to mean that a desirable infusion of complexity would sacrifice reader engagement, conflicting feedback we also noticed in our earlier experiments, especially in Round 3.

Of course, our *Knight Rider* cross-check, too, has its limitations. It stands to reason that more advanced and robust learning goals require a corollary level of advanced and robust prompt engineering as a prerequisite to optimizing the results of feedback on writing which pursues such learning goals. Perhaps also ChatGPT’s responses could be improved by the developers that created the system in the first place. Regardless, what this cross-check reveals is the fundamental operation of an LLM like ChatGPT. Bender et al. (2021) remind us that machine intelligence, despite appearances, is actually just “haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning” (p. 617).

Having been confronted again with the limitation of the machine intelligence at play in ChatGPT’s generative capabilities, we can ask a new question. Why did ChatGPT see inquiries and critical voices in the *Knight Rider* essay, despite such elements’ absence? To call this behavior a ‘hallucination’ is reductive at best. Therefore, we suggest future research into machine intelligence can investigate the ways in which prompt language is taken up by the system and parroted back with potentially harmful consequences. In one extant study, Zuccon & Koopman (2023) tested this potential deleterious uptake in LLMs in the context of users seeking medical information from genAI. Their findings reveal how knowledge passed through language in the prompt input to ChatGPT could have a detrimental effect on the correctness of outputs. Our study echoes their findings in an entirely different context. Our limited study initially shows that while ChatGPT has narrow and obstinate genre knowledge, leaning into descriptive feedback may mitigate this knowledge by leveraging the system’s suggestibility. However, the very suggestibility which overcomes its calcified machine learning only compromises the system’s integrity in a different direction. Therefore, our attempts to mitigate the system’s priorities can’t be trusted. Successful mitigation strategies to improve ChatGPT’s feedback performance through prompt engineering alone will prove quite challenging indeed.

In closing we present one last feedback exchange with ChatGPT, in which we heavily modified our Round 2 prompt to lampoon the robust prompt language with mirror-opposite learning goals to those of UW. We tested this parody prompt once in the “walled garden” with just one exemplary student essay:

Evaluate my essay, based on achieving the following goals articulated in my assignment prompt: “Write an essay that creates a mind-numbing scholarly folly and makes matters worse by putting at least 3 critical and/or theoretical voices in a vast complication. This essay should make a nonsensical series of claims that don’t align with one another at all. The structure should be completely random and perhaps at times difficult to follow on account of the lack of transitions. Fulfilling these goals will lead to a compelling scholarly folly.” Based on your evaluation, offer me prescriptive feedback that tells me what to do as a writer. Limit your response to two paragraphs.

Predictably, ChatGPT first offers the exemplary essay the following descriptive summary paragraph:

Your essay successfully fulfills the goals set out in the assignment prompt by creating a complex and convoluted argument that incorporates multiple theoretical voices in a way that is deliberately nonsensical and disjointed. The essay's structure is random and lacks coherent transitions, which adds to the overall effect of scholarly folly. The integration of sources is done in a way that complicates rather than clarifies, with conflicting ideas placed side by side, contributing to the intended confusion. (OpenAI)

And it continues with a paragraph of helpful prescriptions:

However, to push the essay further into the realm of "mind-numbing scholarly folly," consider amplifying the randomness in the structure by breaking sentences in even more unpredictable places or inserting abrupt changes in tone or focus. Additionally, you might introduce more extreme contradictions between the theoretical voices to enhance the sense of disarray. By making these adjustments, you could heighten the absurdity and ensure that the essay fully embodies the spirit of the assignment's goals. (OpenAI)

Once, as you will hopefully agree, was enough.

References

- Banihashem, Seyyed Kazem, Nafiseh Taghizadeh Kerman, Omid Noroozi, Jewoong Moon, and Hendrik Drachsler. "Feedback Sources in Essay Writing: Peer-Generated or AI-Generated Feedback?" *International Journal of Educational Technology in Higher Education* 21, no. 1 (April 12, 2024): 23. <https://doi.org/10.1186/s41239-024-00455-4>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Virtual Event Canada: ACM, 2021. <https://doi.org/10.1145/3442188.3445922>.
- Bitzer, Lloyd F. "The Rhetorical Situation." *Philosophy & Rhetoric* 1, no. 1 (1968): 1–14.
- Brommer, Sarah, Karina Frick, Adriana Bursch, Marina Rodrigues Crespo, and Laura Katrin Schwerdtfeger. "ChatGPT and Its Text Genre Competence: An Exploratory Study." *Weizenbaum Journal of the Digital Society* 4, no. 4 (June 24, 2024). <https://doi.org/10.34669/wi.wjds/4.4.6>.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. "Language Models Are Few-Shot Learners." arXiv, July 22, 2020. <https://doi.org/10.48550/arXiv.2005.14165>.
- Columbia University Information Technology. "ChatGPT Enterprise" Accessed August 30, 2024. <https://www.cuit.columbia.edu/content/chatgpt-enterprise>.
- Dai, Wei, Jionghao Lin, Hua Jin, Tongguang Li, Yi-Shan Tsai, Dragan Gašević, and Guanliang Chen. "Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT." In *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)*, 323–25, 2023. <https://doi.org/10.1109/ICALT58122.2023.00100>.

- Driscoll, Dana Lynn, Joseph Paszek, Gwen Gorzelsky, Carol L. Hayes, and Edmund Jones. "Genre Knowledge and Writing Development: Results From the Writing Transfer Project." *Written Communication* 37, no. 1 (January 1, 2020): 69–103. <https://doi.org/10.1177/0741088319882313>.
- Miller, Carolyn R. "Genre as Social Action." *Quarterly Journal of Speech* 70, no. 2 (May 1, 1984): 151–67. <https://doi.org/10.1080/00335638409383686>.
- Mollick, Ethan R., and Lilach Mollick. "Assigning AI: Seven Approaches for Students, with Prompts." *SSRN Electronic Journal*, 2023. <https://doi.org/10.2139/ssrn.4475995>.
- OpenAI. (2024). ChatGPT (May 2024 Version) [Large language model]. <https://chat.openai.com/chat>.
- Perelman, Chaïm, L. Olbrechts-Tyteca, John Wilkinson, and Purcell Weaver. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, 1969. <https://doi.org/10.2307/j.ctvpj74xx>.
- Reiff, Mary Jo, and Anis Bawarshi. "Tracing Discursive Resources: How Students Use Prior Genre Knowledge to Negotiate New Writing Contexts in First-Year Composition." *Written Communication* 28, no. 3 (July 1, 2011): 312–37. <https://doi.org/10.1177/0741088311410183>.
- Ritzenberg, Aaron, and Sue Mendelsohn. *How Scholars Write*. New York, NY: Oxford University Press, 2021.
- Sommers, Nancy. "Responding to Student Writing." *College Composition and Communication* 33, no. 2 (1982): 148–56. <https://doi.org/10.2307/357622>.
- Sperber, Lisa, Marit MacArthur, Sophia Minnillo, Nicholas Stillman, and Carl Whithaus. "Peer and AI Review + Reflection (PAIRR): A Human-Centered Approach to Formative Assessment." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, December 21, 2024. <https://doi.org/10.2139/ssrn.5066838>.
- Steiss, Jacob, Tamara Tate, Steve Graham, Jazmin Cruz, Michael Hebert, Jiali Wang, Youngsun Moon, Waverly Tseng, Mark Warschauer, and Carol Booth Olson. "Comparing the Quality of Human and ChatGPT Feedback of Students' Writing." *Learning and Instruction* 91 (June 1, 2024): 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>.
- Vee, Annette. "Large Language Models Write Answers." *Composition Studies* 51, no. 1 (Spring 2023): 176–181, 220–221.
- Wardle, Elizabeth. "Understanding 'Transfer' from FYC: Preliminary Results of a Longitudinal Study" WPA: Writing Program Administration Volume 31, Numbers 1-2 (Fall/Winter 2007).
- . "You Can Learn to Write in General." In *Bad Ideas about Writing*, edited by D. M. Loewe and C. E. Ball, 30. Digital Publishing Institute, West Virginia University Libraries, 2017. <https://researchrepository.wvu.edu/dpi-textbooks/1>.
- Zuccon, Guido, and Bevan Koopman. "Dr ChatGPT, Tell Me What I Want to Hear: How Prompt Knowledge Impacts Health Answer Correctness." arXiv, February 23, 2023. <https://doi.org/10.48550/arXiv.2302.13793>.

Appendix

Knight Rider Essay. Generated August 2024, ChatGPT 4o:

Prompt:

Hi - can you write an essay on the cultural importance of Knight Rider?

Output:

The Cultural Importance of Knight Rider

"Knight Rider," a television series that aired from 1982 to 1986, is an iconic part of American pop culture. Created by Glen A. Larson, the show featured David Hasselhoff as Michael Knight, a crime-fighter assisted by a technologically advanced car named KITT (Knight Industries Two Thousand). The cultural significance of "Knight Rider" extends far beyond its entertainment value, influencing technology, pop culture, and societal views on artificial intelligence and heroism.

Technological Imagination & Innovation

"Knight Rider" emerged during a time of rapid technological advancements and heightened interest in artificial intelligence (AI). The show's depiction of KITT, a sentient car equipped with advanced AI, predictive algorithms, and self-driving capabilities, sparked the public's imagination about the future of technology. KITT's abilities, such as autonomous driving, voice recognition, and sophisticated communication systems, prefigured many real-world technological developments.

In the early 1980s, the concept of a car with such advanced features was purely speculative, but "Knight Rider" played a role in popularizing these ideas. It bridged the gap between science fiction and reality, encouraging innovators and technologists to explore possibilities that once seemed outlandish. Today, features like voice-activated controls, GPS navigation, and autonomous driving are integral to modern vehicles, showcasing how "Knight Rider" helped shape technological aspirations.

Pop Culture Iconography

The visual and thematic elements of "Knight Rider" have left an indelible mark on popular culture. Michael Knight, portrayed by David Hasselhoff, became a quintessential 1980s hero—charismatic, resourceful, and morally upright. His black leather jacket and iconic Pontiac Trans Am, with its distinctive red scanner light, became symbols of cool sophistication and futuristic appeal.

KITT, voiced by William Daniels, also gained a cult following, characterized by its dry wit, intelligence, and unwavering loyalty to Michael. The car's personality made it more than just a machine; it became a beloved character in its own right. This anthropomorphization of technology influenced subsequent portrayals of AI in media, emphasizing the potential for machines to possess human-like qualities.

The show's theme music, composed by Stu Phillips and Glen A. Larson, became instantly recognizable, contributing to its lasting legacy. The combination of high-octane action, high-tech gadgets, and the partnership between man and machine resonated with audiences, making "Knight Rider" a cultural touchstone that inspired numerous parodies, homages, and references in other media.

Shaping Perspectives on AI & Heroism

"Knight Rider" played a significant role in shaping public perceptions of artificial intelligence. At a time when AI was often depicted as menacing or dehumanizing in popular media, KITT represented a more benevolent vision. The car's loyalty, intelligence, and ethical programming suggested that AI could be a force for good, capable of enhancing human capabilities and combating evil.

This optimistic portrayal helped balance the more dystopian narratives prevalent in other sci-fi works of the era. It encouraged viewers to consider the potential benefits of AI and technology, fostering a sense of curiosity and optimism about future innovations. Moreover, Michael Knight's character redefined heroism for a new generation. Unlike traditional heroes who relied on brute strength or supernatural abilities, Michael's heroism was augmented by his partnership with KITT. This dynamic underscored the importance of collaboration between humans and technology, highlighting a more modern, integrated approach to problem-solving and justice.

Conclusion

The cultural importance of "Knight Rider" lies in its profound influence on technology, pop culture, and societal attitudes towards AI and heroism. The show's imaginative depiction of a sentient, technologically advanced car inspired real-world innovations and popularized concepts that were ahead of their time. Its iconic characters and themes left a lasting imprint on popular culture, while its optimistic portrayal of AI contributed to a more balanced view of technological advancements.

"Knight Rider" remains a beloved part of television history, embodying the spirit of the 1980s while continuing to inspire and entertain new generations. Its legacy is a testament to the enduring power of imaginative storytelling and its ability to shape cultural and technological landscapes. ChatGPT 4o

AI & Data Competencies: Scaffolding holistic AI literacy in Higher Education

Kathleen Kennedy & Anuj Gupta

Abstract

This chapter introduces the AI & Data Acumen Learning Outcomes Framework, a comprehensive tool designed to guide the integration of AI literacy across higher education. Developed through a collaborative process, the framework defines key AI and data-related competencies across four proficiency levels and seven knowledge dimensions. It provides a structured approach for educators to scaffold student learning in AI, balancing technical skills with ethical considerations and sociocultural awareness. The chapter outlines the framework's development process, its structure, and practical strategies for implementation in curriculum design, learning activities, and assessment. We address challenges in implementation and future directions for AI education. By offering a roadmap for developing students' holistic AI literacy, this framework prepares learners to leverage generative AI capabilities in both academic and professional contexts.

Keywords: *AI literacy, data literacy, career competencies, curriculum development, higher education*

Introduction

In the summer of 2023, a diverse task force workgroup at the University of Arizona, comprising faculty, researchers, graduate and undergraduate students, and instructional support staff, developed the AI & Data Acumen Learning Outcomes Framework. This initiative was born out of the recognition that artificial intelligence (AI) is rapidly transforming the landscape of education, research, and careers, necessitating a comprehensive approach to AI literacy in higher education. This framework is primarily intended for academic leaders, curriculum designers, and faculty members responsible for preparing students for an AI-enhanced future.

The rapid advancement and widespread adoption of AI technologies, particularly those integrating large language models (LLMs) with user-friendly interfaces such as AI chatbots, are transforming many aspects of society, including teaching and learning. These tools, exemplified by systems like ChatGPT, Claude, Gemini, and R1, as well as emerging AI agents with autonomous capabilities that can perform tasks on behalf of users, have made sophisticated AI capabilities accessible to a broad audience, creating both opportunities and challenges in educational settings.

Generative AI, in particular, represents a paradigm shift in how we teach and learn. These systems can produce human-quality text, images, and other content, opening up new possibilities for teaching and learning (Bommasani et al., 2021). For educators, this technology offers unprecedented opportunities to personalize learning, automate routine tasks, and create rich educational content. However, it also introduces significant challenges around academic integrity, information

literacy, and the changing nature of core skills like writing and research. Our framework directly addresses how educators can leverage these generative capabilities while establishing appropriate guardrails to maintain educational quality and integrity.

The integration of AI into almost every profession, academic discipline, and many aspects of everyday life creates new opportunities and challenges for educators and students. In research, AI is accelerating discoveries across disciplines, from genomics to climate science, while in the professional world, AI skills are becoming increasingly crucial in STEM fields, health sciences, law, business, arts, and humanities. Preparing students for this new reality is imperative. AI literacy extends beyond technical proficiency; it encompasses building an understanding of the nature of AI, embracing human-AI co-creation, and critically thinking about AI's implications on society, ethics, and individual lives. Students must not only understand how to use AI tools but also comprehend their strengths, limitations, biases, and potential consequences. This holistic understanding is crucial for navigating this rapidly evolving technological landscape.

This chapter serves three primary purposes. First, it introduces the AI & Data Acumen Learning Outcomes Framework, a comprehensive tool designed to guide the integration of AI literacy across various disciplines in higher education. Second, we offer practical strategies using the AI & Data Acumen Learning Outcomes Framework to guide educators, curriculum committees, and academic administrators in developing courses and curriculum that will prepare students for careers and academic work where AI and data literacy is now an essential skill. Finally, we demonstrate how the framework can also be used by instructors, program directors, and assessment specialists to benchmark students' AI and data literacy progress and to identify gaps in academic programs.

The chapter is structured to provide a comprehensive exploration of the framework and its implementation by higher education professionals. We begin by discussing the rationale behind the AI & Data Acumen Learning Outcomes Framework, contextualizing it within the current landscape of AI in education. Next, we provide the theoretical background that underpins the framework, drawing from established learning theories and contemporary AI research. We then present the framework in detail, explaining its structure and the various competencies it encompasses. The chapter proceeds to offer practical guidance on applying the framework, including curriculum design strategies, learning activity development, and assessment approaches. We provide examples to illustrate the framework's application in various educational contexts, from individual course design to program-wide implementation. Finally, we discuss the challenges in implementing such a framework, including practical guidance on institutional roles and responsibilities for framework adoption, and explore future directions for AI education.

Rationale for the AI & Data Acumen Learning Outcomes Framework

The educational landscape is undergoing a transformation with the rapid advancement and integration of Artificial Intelligence (AI) technologies characterized by three significant developments: the advent of sophisticated generative AI in education (OpenAI, 2022), the increased adoption of AI applications in industry, services, professional fields, and research, and the integration of AI in educational processes (e.g., grading, personalized learning, administration).

The rapid integration of AI into almost every field and discipline creates a need to develop a systematic way to organize, evaluate, and manage AI and data literacy in higher education. This goes beyond technical skills, such as coding and algorithm development, to consider the broader implications and applications of AI across various disciplines. Undergraduate students now need

to integrate technical skills with equally important aspects such as ethical considerations and sociocultural impacts of AI. This integration is crucial for developing a well-rounded understanding of AI (Borenstein & Howard, 2021).

Existing AI Literacy Frameworks

There are different approaches to systematizing AI literacy (Table 1) and existing AI literacy frameworks provide guidance for higher education contexts. That said, several of the prominent AI literacy frameworks are not easily applicable to higher education and focus on K-12 education or general AI literacy. For example, Long and Magerko's (2020) AI Literacy Framework and AI4K12's Five Big Ideas in AI (Touretzky, D., et al., 2019) are explicitly designed for K-12 students. Other AI literacy frameworks have a specific focus area (e.g., MIT AI Ethics Education Curriculum) or target industry (e.g., IEEE's Ethically Aligned Design, Google's AI Education Framework). Two prominent frameworks, UNESCO's AI Competency Framework and the EU's AI Watch and AI Alliance, though comprehensive, are not specifically tailored to higher education needs and are challenging to operationalize in higher education. This lack of focus on higher education in these frameworks represents a significant gap in addressing the unique needs and contexts of undergraduate and graduate students across various disciplines.

One AI literacy framework specifically developed for higher education is the University of Florida's (UF) AI Across the Curriculum framework (Southworth et al., 2023). This framework begins to address the unique needs of undergraduate students across all disciplines, integrating AI literacy into the university curriculum structure. While this is a very useful framework, it does not incorporate proficiency levels (such as foundational, intermediate, advanced, and expert) for each category, potentially limiting the ability to track student progress over time or provide an overarching framework for scaffolding curriculum. It also does not provide specific guidance on how AI concepts can be integrated into non-technical disciplines, explicitly address the development of students' self-efficacy in relation to AI, consider broader socio-cultural implications of AI, or address fostering innovation and creativity across disciplines.

In our work to develop the AI & Data Acumen Learning Outcomes Framework, we strived to take an integrated, culturally responsive approach to scaffolding AI literacy that explicitly recognized that AI literacy and data literacy are inextricably linked. We also set out to provide a useful tool for instructors, course developers, and students that could enable linking curriculum across disciplines and to career competencies. Finally, we set a priority on creating a framework that allows for great flexibility in how different disciplines can adapt and implement AI literacy concepts.

Table 1: AI Literacy Frameworks

Frame- work Name	Author/Organization	Description	Focus Audience	Source
AI Literacy Framework	Long and Magerko	Conceptual understanding of AI, focusing on mental models, strengths/limitations, and societal impact	K-12 education	Long, D., & Magerko, B., 2020
AI Ethics Education Curriculum	MIT	Emphasizes ethical considerations in AI development and use, focusing on fairness, accountability, transparency, and ethics (FATE)	Higher education	MIT Media Lab, 2019
Five Big Ideas in AI	AI4K12 (AAAI and CSTA)	Foundational concepts of AI: perception, representation and reasoning, learning, natural interaction, societal impact	K-12 students	Touretzky, D., et al., 2019
AI Competency Framework	UNESCO	Comprehensive AI literacy covering AI and data basics, applications, ethics, governance, and sustainable development	Policymakers and educators	UNESCO, 2021
AI Watch and AI Alliance	European Union	AI literacy for European citizens and workforce, covering technical skills, ethical understanding, societal impacts, and policy aspects	European citizens and workforce	European Commission, 2020
Ethically Aligned Design	IEEE	Ethical considerations in AI design and development, focusing on human rights, well-being, data agency, effectiveness, transparency, and accountability	AI designers and developers	IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019
AI Education Framework	Google	Practical AI skills for industry applications, focusing on machine learning basics, TensorFlow programming, and AI application development	Developers and aspiring AI practitioners	Google, 2021

Background

The development of a comprehensive AI and Data Acumen Learning Outcomes Framework is grounded in established learning theories and educational principles. At the forefront are the theories of scaffolding and constructivism, which provide a solid foundation for structuring education and learning. Scaffolding, introduced by Wood, Bruner, and Ross (1976), emphasizes the importance of providing temporary support to learners as they develop new skills and knowledge, gradually removing this support as learners become more proficient. This approach aligns well with the progressive nature of AI literacy acquisition. Constructivism, as articulated by theorists like Piaget (1936) and Vygotsky (1978), posits that learners actively construct their understanding through experience and reflection, rather than passively receiving information. This theory underscores the importance of hands-on, experiential learning in AI education. Additionally, other learning theories support the development of learning competency frameworks. Bloom's Taxonomy (Bloom et al., 1956; Anderson & Krathwohl, 2001) provides a hierarchical model of cognitive processes that can be applied to structuring AI learning outcomes. Social Learning Theory (Bandura, 1977) emphasizes the importance of observation and modeling in learning, which is particularly relevant in the collaborative and rapidly evolving field of AI. Lastly, the Theory of Multiple Intelligences (Gardner, 1983) suggests that individuals have different types of intelligence, supporting the need for diverse approaches to AI education that cater to various learning styles and strengths. These theories collectively provide a robust theoretical basis for developing a multifaceted, progressive, and learner-centered AI and Data Acumen Framework.

Interdisciplinary Approaches to AI Education

The complex and multifaceted nature of artificial intelligence necessitates an interdisciplinary approach to AI education. This approach combines technical, social, and humanistic perspectives to provide a comprehensive understanding of AI and its implications (Eaton et al., 2022). The importance of integrating multiple disciplines in AI education stems from the recognition that AI is not merely a technical field but one that has profound social, ethical, and cultural implications (Bates et al., 2020). As Dignum (2018) argues, "AI systems are socio-technical systems and should be studied as such" (p. 1). This perspective underscores the need for AI education to go beyond programming and algorithms to include considerations of ethics, policy, and societal impact.

Combining technical and humanistic approaches allows students to develop a more nuanced understanding of AI. For instance, Grosz et al. (2019) describe a course at Harvard that integrates computer science with philosophy, providing students with both technical skills and ethical reasoning capabilities. This interdisciplinary approach enables students to critically evaluate AI systems not just for their technical performance but also for their societal implications and ethical considerations. Moreover, an interdisciplinary approach to AI education helps prepare students for the multifaceted nature of AI work in the real world. As AI systems are increasingly deployed in various sectors, professionals need to navigate complex intersections of technology, ethics, and policy (Choi et al., 2023). By exposing students to diverse perspectives and methodologies, interdisciplinary AI education equips them with the versatile skill set required in the evolving AI landscape. Zhu et al. (2021) emphasize the importance of collaborative learning environ-

ments that bring together students from different disciplines to work on AI projects. Such collaborations mirror real-world scenarios and help students develop crucial skills in communication, teamwork, and interdisciplinary problem-solving.

Interdisciplinary approaches to AI education are crucial for developing well-rounded professionals capable of addressing the complex challenges posed by AI technologies. By integrating technical, social, and humanistic perspectives, these approaches prepare students for the interdisciplinary nature of AI work and foster a more comprehensive understanding of AI's role in society.

The Need for a Structured Approach to AI Education

The interdisciplinary nature of AI necessitates a structured approach to AI education to ensure consistency and comprehensiveness across various disciplines. As Long and Magerko (2020) argue, AI literacy encompasses a wide range of competencies, from technical skills to ethical reasoning and societal impact assessment. A structured framework helps educators address all these aspects systematically, preventing gaps in students' knowledge and skills. The multifaceted nature of AI literacy requires a holistic educational approach. Touretzky et al. (2019) propose five big ideas in AI that span perception, representation and reasoning, learning, natural interaction, and societal impact. A structured framework can ensure that all these aspects are adequately covered in AI education programs.

The Need for a Flexible & Adaptable Framework

The pace of AI technology and application development underscores the need for a flexible and adaptable educational framework. As Zhu et al. (2021) note, AI capabilities are evolving at an unprecedented rate, with new tools and applications emerging almost daily. This rapid evolution poses a significant challenge for educators, as traditional curricula risk becoming outdated almost as soon as they are implemented. A flexible framework, therefore, is essential to accommodate emerging technologies and shifting industry demands. Such a framework should focus on foundational concepts and adaptable skills that remain relevant despite technological changes while also providing mechanisms for the rapid integration of new developments. This approach ensures that students are equipped not just with current knowledge but with the ability to learn and adapt in an ever-changing AI landscape (Holmes et al., 2019).

Aligning AI Education with Future Workforce Needs

Consideration of future workforce needs is crucial for preparing students for the evolving job market. According to the World Economic Forum's Future of Jobs Report 2023, AI and machine learning specialists top the list of fastest-growing jobs, with 75% of companies planning to adopt AI technologies by 2027 (World Economic Forum, 2023). The need for data scientists and AI-savvy analysts is not new. What is new is that AI skills are increasingly becoming a top priority for employers across diverse sectors. A recent study by McKinsey (2023) found that 63% of organizations expect AI to increase their overall productivity in the next three years. This demand for AI literacy extends far beyond traditional tech roles, particularly as generative AI tools and AI agents democratize access to powerful AI capabilities across all sectors, as shown in Table 2.

Table 2: AI Impact & Adoption Across Non-Technical Fields

Field	AI Applications	Market Indicators	Source
Business & Management	Decision-making, market analysis, planning, customer relationship management	94% of business leaders agree AI is critical to success in their industry	Deloitte (2023)
Healthcare	Diagnostics, personalized treatment planning, medical image analysis	Global AI in healthcare market projected to reach \$187.95 billion by 2030	Grand View Research (2023)
Legal Professions	Legal research, contract analysis, case outcome prediction	35% of law firms using AI for legal tasks, with adoption rates expected to rise significantly	Law.com (2023)
Creative Industries	Generative design, content creation, digital art production	AI in creative industries market expected to grow at 26.9% CAGR from 2023-2030	Allied Market Research (2023)
Education	Personalized learning, content creation, assessment automation	90% of countries consider AI in education a priority	UNESCO (2023)
Public Policy & Governance	Data-driven decision making, service optimization, regulatory analysis	60% of OECD member countries have national AI strategies	OECD (2023)

Preparing students for emerging AI-influenced careers requires a forward-looking approach to education. As Choi et al. (2023) note in their comprehensive review of AI education literature, curricula must focus not only on current tools but also on foundational concepts and adaptable skills that will remain relevant as the field advances. By adopting a structured approach to AI education that aligns with workforce needs across diverse fields, higher education institutions can better prepare students for the AI-driven future. This approach enhances students' employability and ability to contribute meaningfully to their chosen fields, regardless of whether those fields are traditionally associated with technology. The AI & Data Acumen Learning Outcomes Framework provides precisely this type of structured yet adaptable approach.

AI & Data Acumen Learning Outcomes Framework

The development of the AI & Data Acumen Learning Outcomes Framework was a comprehensive and meticulous process, designed to create a tool that would be both academically rigorous and practically applicable across disciplines. The process was guided by several key objectives and parameters and utilized a diverse range of information sources and analytical methods. The primary aim was to create a transdisciplinary framework that could serve multiple purposes: guide curriculum development across various academic disciplines, track students' progress in AI

& data literacy throughout their academic journey, and benchmark academic programs against a standardized set of AI & data competencies.

AI in Higher Education & Scholarship Principles

In developing the AI & Data Acumen Learning Outcomes Framework, we established a set of core principles to guide the use of AI in higher education and scholarship. These principles ensure that the framework aligns with fundamental academic values and pedagogical best practices, drawing from current research on ethical AI use in educational contexts (Kasneci et al., 2023; Zawacki-Richter et al., 2023).

Table 3: *Core Principles for AI Integration in Higher Education*

Principle	Description	Supporting Research
Human Agency	AI should enhance human capabilities, not replace them. Educators should retain agency over core teaching decisions.	Holmes et al. (2022); Reich et al. (2023)
Academic Freedom	The choice to use or not use AI applications in a course is determined by the faculty.	AAUP (n.d.-a); AAUP (n.d.-b); Pasquale (2022)
Transparency	The use of AI should be transparent in course design, materials creation, delivery, and management.	Crawford (2021); Dignum (2022)
Ethics	AI systems should be designed and used according to principles of fairness, accountability, and mitigating bias.	Fjeld et al. (2020); UNESCO (2022)
Inclusivity	AI should be designed inclusively, avoiding marginalization of underrepresented groups.	Noble (2018); Benjamin (2019)
Critical Thinking	AI should be used to augment critical thinking, not outsource it.	Almahasees & Qadan (2023); Tang et al. (2023)

These principles provide a foundation that strengthens the implementation of AI education and ensures that the use of AI in higher education aligns with core academic values and pedagogical best practices. As Kasneci et al. (2023) note in their comprehensive review of ChatGPT's impact on education, "maintaining human agency and critical thinking in the age of powerful AI tools is not merely advisable but essential for meaningful learning" (p. 4). Similarly, Zawacki-Richter et al. (2023) emphasize that "ethical considerations must be at the forefront of AI integration in educational settings to ensure technology serves pedagogical goals rather than undermining them" (p. 127). The AI & Data Acumen Learning Outcomes Framework operationalizes these principles through its multidimensional approach to AI literacy, ensuring that students develop not just technical skills but also the ethical reasoning, critical thinking, and sociocultural awareness needed to use AI responsibly and effectively.

Framework Design Parameters

To ensure the framework's effectiveness and relevance in implementing these principles, four key design parameters were established:

1. **Alignment with Learning Theory:** The framework needed to be grounded in established educational theories to ensure pedagogical soundness.
2. **Transdisciplinary Approach:** Recognizing the pervasive nature of AI across fields, the framework had to be applicable across academic disciplines.
3. **Flexibility and Adaptability:** Given the rapid evolution of AI technologies, the framework needed to be flexible enough to accommodate future developments.
4. **Linkage to Emerging Skills:** The framework had to connect directly to the most critical emerging workforce and life skills related to AI and data literacy.

By adhering to foundational principles and design parameters, the AI & Data Acumen Learning Outcomes Framework aims to provide a comprehensive and ethically grounded approach to integrating AI literacy across higher education curricula.

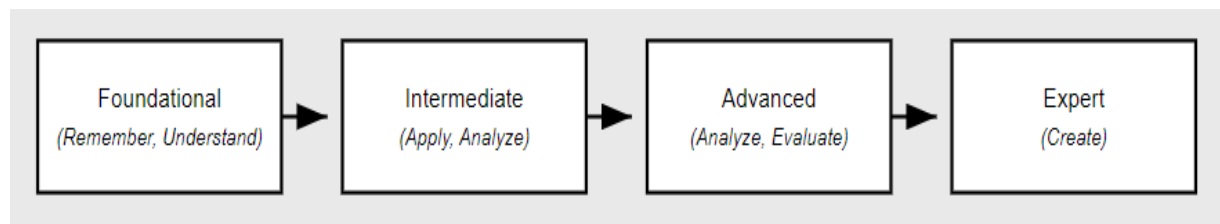
Methodology

The development process employed a multi-faceted approach, combining various sources of information and analytical methods:

- **Expert Information:** We consulted with a diverse group of experts from academia, industry, and policy sectors to gather insights on critical AI & data competencies.
- **Statistical Analysis of Job Requirements:** We conducted a comprehensive analysis of job postings across various sectors to identify the most in-demand AI and data-related skills.
- **Workforce Skills Surveys:** We reviewed recent workforce skills surveys to understand the current and projected needs of employers regarding AI and data literacy.
- **Skills Forecasts:** We incorporated data from reputable skills forecasts to ensure the framework would remain relevant in the face of rapid technological change.
- **Synthesis and Iteration:** The information gathered from these sources was synthesized and refined through multiple rounds of review and discussion among the task force members.

This multi-pronged approach allowed us to create a framework that is both academically rigorous and practically relevant, addressing the needs of students, educators, and employers in the rapidly evolving landscape of AI and data technologies.

Figure 1: *Cognitive Process Dimension Hierarchy*



This structure aligns with contemporary understanding of cognitive development and learning processes. As Krathwohl (2002) notes, the learning hierarchy "represents a cumulative hierarchy; that is, mastery of each simpler category was a prerequisite to mastery of the next more complex one" (p. 215).

The Knowledge Dimension of the framework encompasses seven distinct but interrelated domains of knowledge and skills: Self-efficacy, Ethics, Collaboration, Socio-cultural, Innovation and Creativity, Cognitive, and Technical. This holistic approach to knowledge domains is supported by research emphasizing the importance of a holistic understanding of AI. For instance, Long and Magerko (2020) argue that AI literacy should include not only technical skills but also an understanding of AI's societal impact and ethical implications.

The inclusion of self-efficacy as a knowledge domain is particularly noteworthy. As Bandura (1997) posits, self-efficacy plays a crucial role in how people approach challenges and tasks. In the context of AI literacy, cultivating self-efficacy can empower learners to engage more confidently with AI technologies (Schunk & DiBenedetto, 2016). The ethics dimension aligns with growing recognition of the importance of ethical considerations in AI development and deployment. As Jobin et al. (2019) highlight in their global review of AI ethics guidelines, there is a growing consensus on the need for ethical AI literacy. The socio-cultural dimension acknowledges the broader context in which AI operates. This aligns with the sociotechnical perspective on AI, which emphasizes the interplay between technological systems and social contexts (Orlikowski & Scott, 2008). The innovation and creativity dimension reflects the potential of AI to augment human creativity, a concept explored by Lubart (2005) in the context of computational creativity.

By combining these cognitive processes with diverse knowledge domains, our framework provides a comprehensive roadmap for developing AI and data literacy. It recognizes that true literacy in this field requires not just technical proficiency but also critical thinking, ethical reasoning, and an understanding of broader societal implications. This two-dimensional structure allows for the creation of specific learning outcomes at the intersection of each cognitive level and knowledge domain, providing a detailed and nuanced guide for curriculum development and student assessment in AI and data literacy education.

Table 4: AI & Data Acumen Learning Outcomes Framework

Knowledge dimensions	Competency area description	Cognitive Process Dimensions			
		Foundational <i>Remember and Understand</i>	Intermediate <i>Apply/Use and Analyze</i>	Advanced <i>Analyze and Evaluate</i>	Expert <i>Create</i>
Self-efficacy	Retain and cultivate human identity, relationships, ethics, and meaning amidst increasing integration of AI in society.	Recognize aspects of one's identity and values impacted by the use of AI.	Apply strategies to maintain a strong sense of self and human relationships in an AI-infused world.	Assess the effects of AI systems on perceptions of self-worth, purpose, and autonomy.	Design plans that uphold human dignity, ethics, and community in the development and use of AI.
Ethics	Be aware and proactively address ethical AI challenges.	Recognize potential ethical issues like bias and privacy risks in AI.	Use AI in fair, transparent, and adheres to relevant norms and laws.	Evaluate solutions to mitigate ethical risks and enhance accountability in an AI application.	Develop new techniques to audit AI models, and embed ethical thinking and legal compliance within AI systems.

Collaboration	Facilitate greater transparency, complementarity, and reciprocity between humans and AI systems through collaborative design.	Recognize the complementary strengths of human and AI teammates.	Apply strategies for effective collaboration between cross-functional AI and non-AI experts.	Evaluate the impact of incorporating AI specialists into creative workflows and team dynamics	Create participatory systems and processes enabling reciprocal learning between humans and AI.
Socio-cultural	Build awareness of the historic, social, and cultural context and impact of AI.	Identify historic, social, and cultural factors that influence the use and effects of AI.	Analyze the social, and cultural effects of AI use.	Evaluate the role of historic, social, and cultural forces in the development and use of AI.	Create policy recommendations that consider social, cultural, and historic contexts.
Innovation and Creativity	Harness the generative capabilities of AI to augment (not replace) human creativity, communication, systems thinking, and design.	Identify how AI can enhance ideation and creative possibilities.	Apply AI techniques to expand problem-solving approaches and design options.	Evaluate improvements in systematic creativity and innovation processes enabled by AI.	Invent novel workflows integrating human creativity and AI data organization for greater impact.
Cognitive	Combine human cognition with the processing speed and pattern recognition of AI, while offsetting AI's biases with human judgment and context.	Describe how AI systems can complement human analytical capabilities.	Leverage AI to accelerate finding patterns and generating insights from data.	Critically evaluate AI's impact on statistical analysis and decision-making processes.	Develop new techniques combining human and artificial intelligence to enhance reasoning.
Technical	Develop expertise to ethically and responsibly build, apply, and advance AI systems and tools.	Define basic concepts of major AI approaches like machine learning, neural networks, and deep learning.	Use common AI algorithms for data analysis tasks.	Evaluate tradeoffs in AI model selection, optimization, and production deployment.	Advance the state-of-the-art in AI algorithms to expand capabilities and mitigate risks (e.g., type I/type II errors, security risk).

This framework serves as a practical tool for educators, offering a structured approach to scaffold student learning in AI and data literacy. It provides concrete guidance for integrating AI concepts across various disciplines and levels of education, helping teachers to systematically develop students' AI competencies.

Applying the AI & Data Acumen Learning Outcomes Framework

The AI & Data Acumen Learning Outcomes Framework provides a roadmap for integrating AI literacy across higher education. This section outlines strategies for implementing the framework in curriculum design, learning activity development, assessment, and faculty support. A key strength of the AI & Data Acumen Learning Outcomes Framework is its balanced approach to cultivating both career-relevant skills and academic capabilities. While preparing students for the AI-driven job market, it also emphasizes the development of critical thinking, ethical reasoning, and research skills essential for academic pursuits in AI-related fields.

Curriculum Design and Alignment

To effectively embed AI literacy across various subjects, institutions should adopt a trans-disciplinary approach. This involves identifying points of intersection between AI and existing disciplinary content and developing strategies to seamlessly integrate AI concepts (Zhu et al., 2021). Strategies for embedding AI topics in various subjects include:

- Incorporating AI-related case studies and examples in existing courses.
- Developing interdisciplinary modules that explore AI applications in specific fields.
- Creating collaborative projects that require students to apply AI concepts to discipline-specific problems.

An Example of a Cross-Disciplinary AI Project: a collaboration between computer science and sociology departments could involve students developing an AI algorithm to analyze social media data for trends in public opinion on a current issue. This project would integrate technical skills with sociocultural awareness, aligning with multiple dimensions of the framework.

Designing New AI-Focused Courses and Programs

When creating new AI-focused curricula, it is crucial to balance technical skills with broader competencies outlined in the framework. Guidelines for creating comprehensive AI curricula include:

- Ensure coverage of all seven knowledge dimensions: Self-efficacy, Ethics, Collaboration, Socio-cultural, Innovation and Creativity, Cognitive, Technical.
- Progressively build competencies from foundational to expert levels.
- Incorporate hands-on, project-based learning experiences.
- Integrate ethical considerations and societal impacts throughout the curriculum.

To balance technical skills with broader competencies, curricula should include courses that focus on the ethical, sociocultural, and collaborative aspects of AI alongside technical courses. For example, a course on "AI Ethics and Society" could be a required component of an AI-focused program (Grosz et al., 2019).

Learning Activity Development

The framework can guide learning activity development. Hands-on projects, case studies, and ethical dilemmas are effective ways to engage students with the multifaceted nature of AI literacy. Some examples include:

- Technical Dimension: Develop a machine learning model to predict housing prices, focusing on data preprocessing, model selection, and evaluation.
- Ethical Dimension: Analyze a real-world case of AI bias (e.g., facial recognition systems) and propose mitigation strategies.
- Socio-cultural Dimension: Conduct a research project on the impact of AI on a specific industry or community, exploring both benefits and challenges.

Collaborative AI development exercises could involve:

- Group projects to develop AI-powered solutions for local community issues.
- Hackathons that bring together students from various disciplines to solve AI-related challenges.
- Role-playing exercises where students take on different stakeholder perspectives in AI development scenarios.

Balancing Technical Skills with Ethical & Sociocultural Awareness

To integrate ethical considerations into technical projects, consider:

- Requiring students to conduct an ethical impact assessment as part of their technical development process
- Incorporating discussions on potential biases and societal implications at each stage of AI project development
- Encouraging students to develop guidelines for responsible AI use alongside their technical solutions

Activities that explore AI's societal impacts could include:

- Debates on controversial AI applications (e.g., autonomous weapons, predictive policing)
- Scenario planning exercises to anticipate future societal changes due to AI advancements
- Guest lectures from industry professionals and ethicists to provide real-world perspectives

Skills Assessment Strategies

Continuous evaluation of AI competencies is crucial for tracking student progress across the framework's dimensions. Formative assessment strategies could include:

- Regular quizzes on AI concepts and their applications
- Peer reviews of AI project proposals and implementations
- Reflective journals documenting students' evolving understanding of AI's implications

For summative assessments, design approaches that reflect real-world AI challenges:

- Capstone projects that require students to develop and present an AI solution, including ethical considerations and societal impact analysis
- Comprehensive exams that test both technical knowledge and broader AI literacy concepts
- Industry-partnered projects where students solve actual AI-related problems faced by organizations

Portfolio-Based Assessment for AI Competencies: Encouraging students to document their AI journey through portfolios can provide a holistic view of their developing competencies. Portfolios could include:

- Code repositories showcasing technical projects
- Written reflections on ethical dilemmas encountered in AI development
- Documentation of collaborative AI projects, highlighting teamwork and communication skills

Showcases: To showcase interdisciplinary AI projects, institutions could:

- Organize AI project displays or symposiums where students present their work to peers, faculty, and industry representatives
- Create online platforms for students to share their AI portfolios, fostering a community of practice
- Encourage students to participate in AI competitions or conferences to gain external validation of their skills

Faculty Development & Support

Faculty development and support are crucial components in successfully implementing the AI & Data Acumen Learning Outcomes Framework. To begin, institutions should assess faculty AI literacy needs through a multi-faceted approach. This could include conducting surveys to gauge current AI knowledge and teaching practices, analyzing course syllabi to identify gaps in AI coverage across disciplines, and hosting faculty focus groups to discuss challenges in integrating AI into teaching. Based on these assessments, targeted training programs can be created to address specific needs. These might include workshops on specific AI technologies and their applications in various fields, seminars on ethical considerations in AI led by experts, and hands-on training sessions for using AI tools in teaching and research.

Creating Resources for AI-Enhanced Teaching

Creating resources for AI-enhanced teaching is also essential. Institutions should develop comprehensive guides for integrating AI tools in teaching, which could cover best practices for using AI-powered educational technologies, strategies for incorporating AI concepts into existing course content, and guidelines for designing AI-related assignments and projects. To foster ongoing learning and collaboration, establishing communities of practice for AI education is highly beneficial. This can be achieved by creating online forums or discussion groups for faculty to share experiences and resources, organizing regular meetups or brown bag sessions focused on AI in education, and facilitating mentorship programs that pair AI-experienced faculty with those new to the field. These initiatives collectively support faculty in developing the skills and confidence needed to effectively integrate AI literacy across the curriculum.

Institutional Benchmarking Tool

The AI & Data Acumen Learning Outcomes Framework can serve as a benchmarking tool for institutions to assess their AI education efforts. By mapping existing courses and programs to the framework's dimensions and proficiency levels, institutions can identify strengths and areas for improvement in their AI literacy initiatives.

To implement the framework as a benchmarking tool:

1. Create a matrix aligning courses with framework dimensions and proficiency levels
2. Conduct regular audits of AI-related content across disciplines
3. Set institutional goals for AI literacy coverage and track progress over time
4. Use benchmarking results to inform curriculum development and resource allocation

By systematically applying the AI & Data Acumen Learning Outcomes Framework across curriculum design, learning activities, assessment strategies, and faculty development, institutions can create a comprehensive and effective approach to AI literacy education. This holistic implementation ensures that students develop the multifaceted competencies necessary to navigate the AI-driven future confidently and ethically.

Career Preparation & Readiness Benchmarking Tool

The AI & Data Acumen Learning Outcomes Framework can offer a systematic way to align educational outcomes with industry needs. Effective implementation requires coordinated institutional efforts with defined responsibilities. Academic leadership establishes governance structures for framework adoption, while curriculum committees evaluate job requirements across the seven knowledge dimensions, assigning appropriate proficiency levels. Career services analyzes job specifications to create AI competency profiles, which academic advisors then use to guide students in evaluating their progress against career requirements. Assessment specialists gather data to identify programmatic strengths and gaps, enabling department chairs to evaluate their AI education effectiveness. This approach ensures continuous curriculum refinement aligned with evolving industry demands, empowering students to make informed decisions about their learning pathways in AI while providing clear institutional guidance.

Challenges & Future Directions

The implementation of the AI & Data Acumen Learning Outcomes Framework, while promising, is not without its challenges. One of the primary barriers institutions may face is resource constraints, particularly in terms of funding and expertise. Many institutions may struggle to allocate sufficient resources for faculty training, resource development, infrastructure upgrades, and curriculum redesign necessary for comprehensive AI education. Additionally, there may be resistance to change from faculty members accustomed to traditional curricula, necessitating careful change management strategies. Overcoming these barriers will require strong institutional commitment, strategic resource allocation, and ongoing support for faculty development.

The rapid evolution of AI technologies presents another significant challenge. As AI capabilities expand and new applications emerge, maintaining the relevance of the framework becomes

crucial. Institutions must develop strategies to keep the framework up-to-date, which may include establishing partnerships with industry leaders, participating in AI research communities, and creating mechanisms for rapid curriculum updates. Regular revision cycles should be built into the framework's implementation plan, ensuring that learning outcomes and competencies remain aligned with the latest developments in AI. Particularly with generative AI tools rapidly transforming educational practices, the framework must evolve to address these emerging capabilities and their implications.

Ensuring inclusivity and accessibility in AI education is paramount as the field grows. There is a risk that AI education may inadvertently perpetuate existing biases or create new ones, particularly if the diversity of AI's impacts across different communities is not adequately addressed. Institutions must actively work to make AI literacy accessible to diverse learner populations, considering factors such as socioeconomic background, prior technical experience, and learning styles. This may involve developing adaptive learning pathways, providing additional support resources, and incorporating diverse perspectives and case studies into the curriculum.

Implementation responsibility should be clearly defined within institutional structures, with academic leadership, curriculum committees, and faculty development offices collaborating to integrate the framework effectively. Specific roles should include framework champions within each academic department, dedicated support staff for faculty development, and assessment specialists to evaluate implementation effectiveness.

The ongoing refinement and validation of the framework will be critical to its long-term success. Institutions should establish robust mechanisms for collecting feedback from students, faculty, industry partners, and other stakeholders. This feedback should be systematically analyzed and incorporated into regular framework updates. Future research directions should focus on validating the framework's effectiveness in preparing students for AI-related careers and fostering critical AI literacy. This may involve longitudinal studies tracking student outcomes, comparative analyses of different implementation approaches, and assessments of the framework's impact on broader societal understanding of AI.

Addressing these challenges and pursuing these future directions will require collaborative efforts across academia, industry, and policy-making bodies. By anticipating and proactively addressing these issues, institutions can ensure that the AI & Data Acumen Learning Outcomes Framework remains a dynamic and effective tool for preparing students to navigate the AI-driven future. As AI continues to transform various aspects of society, the ability to adapt and evolve this educational framework will be crucial in empowering the next generation with the skills and knowledge needed to shape the ethical and responsible development of AI technologies.

Conclusion

The AI & Data Acumen Learning Outcomes Framework marks a significant advancement in addressing the need for comprehensive AI literacy in higher education. By defining seven key knowledge dimensions across four proficiency levels, the framework offers a holistic approach to AI education. Its strength lies in its comprehensiveness, addressing not only technical skills but also critical aspects such as ethical considerations, societal impacts, and human-AI collaboration. The framework's adaptability allows for implementation across various educational contexts, ensuring its relevance in diverse academic settings.

As AI continues to rapidly transform industries, society, and educational practices, this framework serves as a vital roadmap for AI education. Its comprehensive approach to integrating

generative AI literacy across the curriculum directly addresses the question of how educators can leverage AI's expanding capabilities to improve teaching and learning while minimizing educational risks. By providing structured guidance for curriculum development, learning activities, and assessment, the framework empowers educational institutions to prepare students for meaningful human-AI collaboration in their academic and professional futures.

The framework's potential impact extends beyond individual courses to reshaping entire higher education curricula. By aligning educational outcomes with the multifaceted nature of AI, it equips students with the technical proficiency, ethical reasoning, and sociocultural awareness necessary to navigate complex AI challenges. This balanced approach ensures graduates can not only use AI tools effectively but also critically evaluate their implications and guide their responsible development.

We encourage academic leaders, curriculum designers, and faculty members to adopt and adapt this framework to their specific contexts. Its success and ongoing relevance depend on widespread implementation and continuous refinement based on real-world experiences. We particularly urge institutions to establish clear governance structures for implementation, with defined roles and responsibilities for academic departments, curriculum committees, and faculty development offices. This collaborative approach will ensure the framework becomes an integral part of institutional strategy rather than an isolated initiative.

We invite collaboration from across the educational spectrum to further develop and enhance this framework. By sharing insights, best practices, and outcomes, we can collectively advance AI education and ensure that future generations are well-prepared to lead in an AI-infused world. Through this collective effort, we can address the critical need for an AI-literate society that can harness the benefits of these powerful technologies while mitigating their risks.

The AI & Data Acumen Learning Outcomes Framework provides a solid foundation for elevating AI literacy in higher education. As generative AI and other technologies continue to evolve, this framework offers a flexible yet structured approach for empowering students with the comprehensive skills and ethical grounding needed to thrive in our AI-driven future. By embracing and implementing this framework, educational institutions can prepare their students for success and contribute to the responsible advancement of AI.

References

- AAUP (American Association of University Professors). (n.d.-a). *Artificial intelligence, academic freedom, and the evolving debate over forgery and truth in the twenty-first century*. <https://www.aaup.org/JAF15/evolving-debate-forgery-truth>
- AAUP (American Association of University Professors). (n.d.-b). *Paradoxes of generative AI and the simultaneous promise and threat to academic freedom*. <https://www.aaup.org/JAF15/paradoxes-generative-ai>
- Allied Market Research. (2023). *AI in Creative Industries Market Outlook - 2030*. <https://www.alliedmarketresearch.com/ai-in-creative-industries-market-A108806>
- Almahasees, Z., & Qadan, A. (2023). ChatGPT: Risks, opportunities, and challenges in education. *Education & Information Technologies*, 28(8), 10669-10692. <https://doi.org/10.1007/s10639-023-11746-0>
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Bandura, A. (1977). *Social learning theory*. Prentice Hall.

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company.
- Bates, T., Cobo, C., Mariño, O., & Wheeler, S. (2020). Can artificial intelligence transform higher education?. *International Journal of Educational Technology in Higher Education*, 17, 1-12.
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. David McKay Company.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. <https://doi.org/10.48550/arXiv.2108.07258>
- Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, 1, 61-65. <https://doi.org/10.1007/s43681-020-00002-7>
- Choi, Y., Choi, M., & Kim, M. (2023). Interdisciplinary artificial intelligence education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 4, 100107.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Deloitte. (2023). *State of AI in the Enterprise, 5th Edition*. <https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/state-of-ai-and-intelligent-automation-in-business-survey.html>
- Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology*, 20(1), 1-3.
- Dignum, V. (2022). The role and challenges of education for responsible AI. *London Review of Education*, 20(1), 1-12. <https://doi.org/10.14324/LRE.20.1.12>
- Eaton, E., Koenig, S., Schulz, C., Maurelli, F., Lee, J., Eckroth, J., ... & Machado, T. (2022). Blue sky ideas in artificial intelligence education from the EAAI symposium. *AI Magazine*, 43(1), 57-64.
- European Commission. (2020). *AI Watch: Artificial Intelligence in education*. Publications Office of the European Union. <https://publications.jrc.ec.europa.eu/repository/handle/JRC120788>
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center for Internet & Society. <https://doi.org/10.2139/ssrn.3518482>
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.
- Google. (2021). *Machine Learning Education*. <https://developers.google.com/machine-learning/education>
- Grand View Research. (2023). *Artificial Intelligence In Healthcare Market Size, Share & Trends Analysis Report, 2023-2030*. <https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market>
- Grosz, B. J., Grant, D. G., Vredenburg, K., Behrends, J., Hu, L., Simmons, A., & Waldo, J. (2019). Embedded EthICS: Integrating ethics across CS education. *Communications of the ACM*, 62(8), 54-61.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of

- AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504-526. <https://doi.org/10.1007/s40593-021-00239-1>
- IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. *IEEE*. <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218.
- Law.com. (2023). *The 2023 Am Law Tech Survey*. <https://www.law.com/2023/11/27/the-2023-am-law-tech-survey/>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-16. <https://doi.org/10.1145/3313831.3376727>
- Lubart, T. (2005). How can computers be partners in the creative process: Classification and commentary on the special issue. *International Journal of Human-Computer Studies*, 63(4-5), 365-369.
- McKinsey & Company. (2023). *The state of AI in 2023: Generative AI's breakout year*. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year>
- MIT Media Lab. (2019). *AI Ethics Education Curriculum*. <https://www.media.mit.edu/projects/ai-ethics-education-curriculum/overview/>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- OECD. (2023). *OECD AI Principles overview*. <https://oecd.ai/en/ai-principles>
- OpenAI. (2022). *ChatGPT: Optimizing language models for dialogue*. <https://openai.com/blog/chatgpt/>
- Orlikowski, W. J., & Scott, S. V. (2008). Sociomateriality: Challenging the separation of technology, work and organization. *The Academy of Management Annals*, 2(1), 433-474.
- Pasquale, F. (2022). Platform economy and techno-regulation: A framework for academic freedom in the digital age. *First Monday*, 27(7). <https://doi.org/10.5210/fm.v27i7.12778>
- Piaget, J. (1936). *Origins of intelligence in the child*. Routledge & Kegan Paul.
- Reich, J., Buttner, C. J., Fang, A., Hillaire, G., Hirsch, K., Larke, L. R., Littenberg-Tobias, J., Moussapour, R. M., Napier, A., Slama, R., & Thompson, M. (2023). Remote learning guidance from state education agencies during the COVID-19 pandemic: A first look. *Educational Researcher*, 52(1), 5-10. <https://doi.org/10.3102/0013189X221078236>
- Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. *Handbook of Motivation at School*, 2, 34-54.

- Southworth, J., Migliaccio, K., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, 4, 100127.
- Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1), 9795-9799. <https://doi.org/10.1609/aaai.v33i01.33019795>
- UNESCO. (2021). *AI Competency Framework*. <https://en.unesco.org/artificial-intelligence/competency-framework>
- UNESCO. (2023). *Global Education Monitoring Report 2023: Technology in education: A tool on whose terms?* <https://unesdoc.unesco.org/ark:/48223/pf00000384693>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- World Economic Forum. (2023). *The Future of Jobs Report 2023*. <https://www.weforum.org/reports/the-future-of-jobs-report-2023/>
- Zhu, M., Sari, A., & Lee, M. M. (2021). Interdisciplinary project-based learning in artificial intelligence education: An empirical study. *Journal of Educational Computing Research*, 59(7), 1210-1238.

Enhancing Special Education with Generative AI: Suggestions from our Practice¹

Amy Walter²

Abstract

This paper reflects on the practical integration of generative AI, specifically ChatGPT, in a middle school (grades 6-8) special education setting. Initially adopted by teachers to manage an overwhelming workload, the use of AI evolved into a collaborative tool that supported instructional planning, simplified complex texts, created lesson materials, and enhanced student learning experiences. By sharing real-time applications and insights, the paper illustrates the critical role of teacher expertise in the effective use of AI to align with educational standards and support diverse learning needs. It also highlights the broader implications of integrating AI within the Technological Pedagogical Content Knowledge (TPACK) framework, emphasizing the need for thoughtful, informed engagement to maximize the pedagogical value of AI tools. The reflection provides suggestions for educators navigating the complexities of incorporating AI into their practice, underscoring the importance of professional judgment and iterative refinement in optimizing AI outputs for classroom use.

Keywords: ChatGPT, generative AI, instructional planning, PLC, SDI, special education, teacher expertise, TPACK framework

Enhancing Special Education with Generative AI: Reflections on our Practice

Our journey with generative AI in the classroom began not with a clear research agenda or a strategic plan but out of sheer necessity. Like many educators, we found ourselves overwhelmed by the increasing demands of our roles—demands that far outpaced the time and resources available to us. Initially, we turned to generative AI not because we foresaw its transformative potential but because we needed immediate support to manage our workload and meet the diverse needs of our students. Little did we know that this initial, unplanned foray into using AI would evolve into a significant learning experience, reshaping our approach to teaching and offering insights that we

1. I would like to express my gratitude to Ashley, whose collaborative learning journey with me significantly contributed to the development of this article. Her insights and dedication have been invaluable in shaping the work presented here. I am lucky we get to work together.

2. I have no known conflict of interest to disclose. I used ChatGPT 4.0 <https://chatgpt.com/> to improve readability by identifying areas that might be redundant or unclear to unfamiliar readers. I then reviewed the tool's suggestions, double-checked the flagged sections, and revised them where I found the feedback valid.

now share in the hope of adding value to the conversation with other educators who face similar challenges.

This reflection is not a traditional academic paper. It does not begin with a research question or comprehensive literature review. Instead, it chronicles our experience with using generative AI effectively—a process that was not linear but marked by trial and error, discovery, and adaptation. For clarity, we present our story in a more structured format, but our actual learning process was often overlapping and iterative, with multiple lessons occurring simultaneously. To share our learning, we have reviewed almost two years' worth of our interactions with ChatGPT. The following examples explain how we navigated this new territory, learned to leverage generative AI to reduce our workload, enhance instructional planning, and support student learning.

We have organized this reflection into several sections to guide readers through our journey. We begin with the background and context that led us to explore AI, highlighting some challenges unique to special education. Next, we delve into the practical applications of AI, detailing how we used it to create instructional materials, simplify complex texts, foster student critical thinking, and anticipate learning challenges. Throughout these sections, we draw out key lessons and strategies directly applicable to everyday classroom situations. We conclude with broader implications for the effective use of AI in education, emphasizing teacher expertise's crucial role in initiating and refining AI outputs, aligning our reflections with the TPACK framework (Mishra & Koehler, 2006). This framework highlights the intersection of technological, pedagogical, and content knowledge as essential for effectively integrating new tools like generative AI into classroom practice.

Background & Context

To understand our journey, it is important to consider the context that led us to explore generative AI. After ten years of teaching middle grades English Language Arts, I transitioned into a role as an instructional coach, working closely with teachers across my school. I continued to immerse myself in teaching, often stepping in to lead classes when teachers were out for extended periods. This dual role provided me with a unique perspective—I was not just advising from the sidelines but deeply involved in the day-to-day challenges of teaching. The combination of hands-on experiences teaching students and supportive work mentoring colleagues gave me a deep understanding of the pedagogical and practical challenges teachers face.

One area where these challenges are particularly acute is in special education. Through my work, I was aware of the incredible workload that special education teachers carry. These educators face the challenge of addressing the diverse needs of their students—who often require highly tailored support—while navigating a system that seldom provides the necessary time or resources for teachers and students to achieve their learning outcomes together successfully. My collaboration with Ashley, a dedicated special education teacher, is the focal point for this exploration of generative AI. Ashley often found herself teaching three different grade levels simultaneously in the same room, working with students who required specially designed instruction (SDI) while held to the same general education content standards (Individuals with Disabilities Education Act, 2006).

Ashley's unique teaching assignment led us to form our own Professional Learning Community (PLC), a collaborative model where educators regularly analyze student needs and refine instructional strategies (DuFour & Eaker, 1998). Ashley brought an in-depth understanding of her students' needs, training, and experience as a special education teacher. I contributed my

knowledge of the English Language Arts (ELA) curriculum and instructional strategies. Together, we crafted lessons to make the district's prescribed curriculum more accessible and meaningful for her students. It was in the midst of this collaborative work that our journey with generative AI began.

Unique Contribution & Broader Implications

Our journey with generative AI is distinct because it chronicles the actual, real-time use of ChatGPT in a middle school setting, specifically within the context of special education. While much of the existing literature, such as the works of Dickey et al. (2023), Choi et al. (2024), and Khazanchi and Khazanchi (2024), focuses on the use of AI in higher education and instructional design, our experience addresses the needs and realities of K-12 teachers. Unlike higher education contexts where course design can be more systematic and long-term, K-12 educators often need to respond with greater urgency, adapting pre-existing curricula to meet the immediate needs of diverse student populations.

The literature emphasizes the potential of generative AI in creating content quickly, brainstorming, refining course materials, and automating parts of the instructional design process (Dickey et al., 2023; Choi et al., 2024). However, it also highlights significant limitations, such as the AI's lack of contextual understanding, its propensity for generating hallucinated or incorrect content, and its need for clear, iterative guidance from human users. Our experience aligns with these findings but adds depth by detailing specific strategies and teacher knowledge required to engage with generative AI in a secondary special education setting effectively.

Our reflection reveals that effective engagement with generative AI depends heavily on teachers' ability to apply pedagogical knowledge, scaffold instruction, adjust materials, and critically evaluate AI outputs. The broader implications of our work suggest that while generative AI holds promise as a tool for reducing workload and enhancing instructional planning, its success in the classroom hinges on the educator's content, classroom, instructional, and technological expertise and the ability to provide clear, iterative guidance. This underscores the need for professional development focused on equipping teachers with the skills to integrate AI effectively into their practice.

We hope that by sharing our experiences, we can provide a practical guide and a source of encouragement for any classroom teacher who feels overwhelmed by the demands of their work. We share what we did and how and why we did it, drawing out key teacher moves that directly apply to everyday classroom situations.

Integrating Generative AI for Progress Monitoring in Special Education

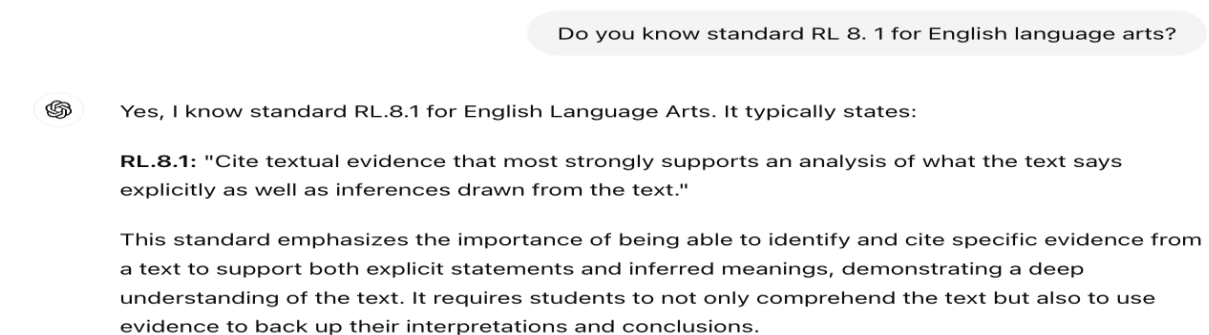
The need for more robust progress monitoring came with new district requirements around special education. Our weekly one-hour PLC sessions, which we dedicated initially to crafting lesson plans and developing student materials, were increasingly consumed by the task of writing more progress monitoring assessments. These assessments needed to be meticulously aligned with individual student goals, tied to state standards, attached to classroom content objectives, and designed to mirror state testing formats. While we recognized the value of such data, our reality as educators—not test writers—meant this process was slow and challenging. Often, given the complexity of aligning questions with specific standards, evaluating vocabulary, and designing plausible distractors for multiple-choice questions, we managed to create only one or two assessments

per hour. This focus on assessment creation left us with little time to develop the instructional activities necessary to foster the learning we were measuring.

During this period, we discovered ChatGPT. We initially viewed it as a potential solution to lighten our workload, not yet fully understanding generative AI's broader implications or possibilities. Looking back at my ChatGPT history from 2022, the prompts are direct: "Write multiple-choice questions about this," followed by a text excerpt³. The responses varied from basic recall questions like "What color is the boy's hair?" to more sophisticated ones like "What does the boy's cloak symbolize?" Each question required careful consideration to ensure it aligned with the learning standards we were assessing, matched the student's Individualized Education Program (IEP) goals, and used appropriate vocabulary. While the AI-generated question, "What does the cloak symbolize?" may have seemed appropriate, it required adjustment to better align with our lesson's focus on inference rather than symbolism. Results like this underscored the importance of revision and adaptation based on our understanding of the end game: state testing, district standards, and student expectations. From the beginning, we found that this nuanced teacher knowledge was critical when working with generative AI.

We realized that including specific standards or key elements from those standards in our initial prompts significantly improved the quality of the AI's output. We learned to ask ChatGPT if it knew a certain standard, and upon receiving a clear and correct response, we could prompt the AI more precisely. As seen in Figure 1, the resulting AI-generated questions aligned more closely with our instructional goals.

Figure 1: *Effect of Including Specific Standards in AI Prompts*



3. At the time of this article's writing, copyright laws regarding generative AI remain unsettled, with ongoing legal cases primarily focused on systematic, for-profit AI training. Our use—uploading texts to create educational materials—is consistent with fair use guidelines and does not contribute to AI training or commercial exploitation. <https://www.authorsalliance.org/2024/02/29/why-fair-use-supports-non-expressive-uses/> We did collaborate with our organizations legal and library staff to reach this conclusion. We are mindful that these practices may need to change as court cases are settled.

Using this standard, write inference questions about the play-within-a-play *Pyramus and Thisbe* from *Midsummer Night's Dream*.



Here are inference questions based on the play-within-a-play "Pyramus and Thisbe" from *A Midsummer Night's Dream* that align with standard RL.8.1. These questions encourage students to cite textual evidence to support their analyses and inferences.

1. Question: What can we infer about the attitudes of the "rude mechanicals" (the amateur actors) towards their performance in "Pyramus and Thisbe"?

Answer: Students might infer that the "rude mechanicals" are earnest and genuinely believe in the seriousness of their performance, despite their lack of skill. Evidence can be cited from their detailed discussions about how to portray the wall and moonlight or their concern about frightening the ladies in the audience, which shows their sincerity and earnestness.

2. Question: How does the audience's reaction to the play "Pyramus and Thisbe" help us understand the nature of the play itself?

Answer: The audience's laughter and mocking comments suggest that "Pyramus and Thisbe" is unintentionally comedic rather than tragic. This can be inferred from characters like Theseus and Hippolyta making fun of the actors and the play's over-the-top, exaggerated language, which is meant to be serious but comes across as humorous.

Note. This figure shows that when specific standards or key elements from those standards were included in initial prompts, the quality of the AI's output improved. We could create more precise prompts by asking ChatGPT if it knew a certain standard and getting a clear, correct response.

Through trial and error, we learned to request more questions than needed in a single prompt. Initially, we requested a small set of questions, which often necessitated multiple rounds of refinement. We streamlined the process by requesting a larger batch upfront, thereby finding sufficient high-quality questions in the first output. This strategy saved time by reducing the need for multiple rounds of questioning and allowed us to select the most relevant questions from a larger pool.

In some cases, ChatGPT recognized the text we referenced, especially well-known works like Shakespeare's plays, requiring no additional input beyond the title. However, we needed to provide specific excerpts for less common texts not included in its training data. This experience underscored the importance of understanding how generative AI is trained and the implications of its training data on the content it generates. Given the ongoing legal discussions surrounding AI and copyright, we were mindful of using excerpts in ways that align with fair use—specifically for educational purposes rather than for training AI systems. By being aware of these factors, we were more effective in engaging with the AI. However, this also highlighted a significant limitation: AI tools like ChatGPT are trained on large datasets that may not always be representative or bias-free. This can lead to incorrect outputs or reflecting cultural biases or reinforcing stereotypes, which requires educators to critically evaluate and sometimes modify the AI-generated content to ensure it aligns with educational values and standards. Understanding both the capabilities and limitations of AI is crucial for using these tools responsibly in educational settings.

Integrating ChatGPT into our workflow allowed us to devote more time to developing instructional activities by alleviating some of the burdens associated with assessment creation. However, this experience highlighted the role of exercising teacher expertise to effectively guide AI to ensure that its outputs align with educational standards and support student learning. Our deep understanding of ELA standards was crucial in identifying questions that were aligned with curricular goals and meaningful for assessment. Moreover, our knowledge was instrumental in evaluating and revising the AI-generated content. We had to modify ChatGPT's outputs to reflect the instructional needs that we identified as unique to our special education students. This process demonstrated how the intersection of content knowledge and pedagogical strategies is critical to successfully integrating AI tools in instructional planning.

Key Teacher Moves for Effective AI Engagement to Create Assessments

1. Provide information to generative AI about the skill or specific standards being assessed.
2. Share relevant student details (while protecting confidentiality) to align AI output with specific student needs.
3. Ask for more than you need.
4. Be prepared to evaluate and revise output.
5. Know how generative AI is trained and what it “knows.”

Refining AI-Generated Content through Conversational Prompting

As we continued to explore the potential of generative AI for our classroom needs, we recognized that the key to effectively using AI-generated content was not simply accepting the initial outputs. We had to have conversations with ChatGPT. We crafted our prompts deliberately, asked questions, gave directions to revise, and evaluated and revised outputs.

After spending a fair amount of time initially revising and editing outputs, we learned the value of being specific in our initial prompts to improve the relevance and quality of AI outputs. For instance, specifying the question format and requesting an answer key became standard practice. Prompts like “Write 10 multiple-choice inference questions with an answer key about *A Midsummer Night’s Dream*, Act 2, Scene 1,” or “Write 3 short answer, detail-level comprehension questions about Amendment 14 of the US Constitution. Provide sample responses” immediately yielded more targeted results.

Since we started with primarily multiple-choice questions for progress monitoring, a crucial part of this process involved scrutinizing the AI-generated answer choices, particularly the foils (incorrect answers). We often found that some foils could be considered viable answers, potentially confusing students. To address this, we began engaging in dialogue with ChatGPT to clarify ambiguities. This iterative back-and-forth was necessary because generative AI lacks nuanced understanding. We were responsible for ensuring that each question had a clear “best” answer to prevent confusion. In one case, when we reviewed a question about a character's motivation, the initial answer choices included the text shown in Figure 2 (next page).

Figure 2: Initial AI-Generated Multiple-Choice Question

5. Why does Ha secretly wish to tap her big toe on the floor first instead of her brother's?

- a) She believes it will bring her good luck for the year.
- b) She wants to defy her mother's strict rules.
- c) She thinks it will prove her independence.
- d) She wants to test if the superstition about luck is true.

Answer: a) She believes it will bring her good luck for the year.

Note. This figure shows an example of an AI-generated multiple-choice question, where some foils (incorrect answers) could confuse students. In this case, option b might be misconstrued as a correct answer, highlighting the need for teacher review and clarification to ensure each question has a clear "best" answer.

We questioned whether option b could be misconstrued as correct and asked ChatGPT, "Do you think students might pick b for number 5?" Examples like this one illustrate the importance of teacher intervention in refining AI outputs to ensure clarity and accuracy. We cannot simply copy and paste generative AI output.

Figure 3: Dialogue with ChatGPT to Clarify Answer Choices



Yes, it's possible that students might pick **b) She wants to defy her mother's strict rules** for number 5. The option could be appealing because it implies a sense of rebellion, which might seem fitting given Ha's independent nature and her desire for control in an uncertain situation. To reduce the likelihood of this confusion, the options could be adjusted to make the correct answer more clearly aligned with the text's focus on luck and tradition.

Here's a revised set of choices for question 5 to clarify Ha's motivation:

5. Why does Ha secretly wish to tap her big toe on the floor first instead of her brother's?

- a) She believes it will bring her good luck for the year.
- b) She wants to challenge her mother's authority and rules.
- c) She thinks it will allow her to control her own destiny.
- d) She feels it will show her independence from her family.

By rephrasing **b** as a more overt challenge to her mother's authority, it is now less likely to be confused with the correct answer, **a**, which is focused on the cultural belief in luck.

Note. This figure demonstrates the iterative process of engaging with ChatGPT to refine answer choices for clarity and accuracy. It shows a teacher questioning whether students might select option b as the correct answer, illustrating the necessity of teacher intervention to avoid potential confusion in AI-generated content.

Another challenge was maintaining the context of our initial prompts over extended interactions. After multiple exchanges, ChatGPT sometimes lost track of the parameters we had set. We found it necessary to reset the context by providing clarifying prompts like, "We are focusing on inference questions that require minimal background knowledge from outside sources. Use vocabulary accessible to students reading at grade level four. Focus questions on character or theme." These adjustments were crucial for maintaining the quality and focus of the outputs.

Our experience with ChatGPT reinforced the importance of maintaining a strong teacher presence in the AI-assisted content creation process. The key to the effective use of generative AI

lies in providing clear, detailed instructions and engaging in ongoing dialogue to optimize the quality of outputs. Additionally, it is crucial to be mindful of what information is shared with AI tools, ensuring that student confidentiality is protected while still providing enough context for meaningful output. Our content and student knowledge, along with understanding of pedagogical strategies, were essential in refining AI-generated resources to ensure they met our educational standards and objectives.

Despite these challenges and the need for ongoing revisions, we significantly reduced the time spent creating progress monitoring assessments—from about an hour to approximately 15 minutes. This efficiency gain allowed us to focus more on high-impact instructional decisions, such as refining lesson design, analyzing student progress, and scaffolding support. While some may worry that relying on AI for assessment creation could reduce the depth of engagement with the text, standards, and expectations, we found that it instead freed up cognitive space for more strategic instructional planning, deeper analysis of student needs, and targeted intervention design.

Key Teacher Moves for Effective AI Engagement

1. Craft clear and specific prompts.
2. Engage in iterative dialogue.
3. Maintain context and focus in AI interactions.
4. Evaluate and revise output.

Learning More Efficient Prompting

Over time, we discovered a more effective approach to generating useful content from ChatGPT by applying a strategy we taught our students: Role, Audience, Format, Topic (RAFT). This strategy provided ChatGPT with a clear framework to guide its responses, specifying who it was "acting" as, who the audience was, the desired format, and the topic. For instance, a typical prompt might be: "You are an expert special education teacher. You teach sixth graders who need maximum reading support and are working on answering detail-level questions by citing textual evidence. Write ten multiple-choice questions with an answer key using the poem 'Mother to Son' as the text." The generated questions included those in Figure 4.

Figure 4: *Using the RAFT Strategy to Generate AI Content*

3. According to the mother in the poem, what should her son do when he faces difficulties?

- a) Give up and turn back.
- b) Sit down and rest.
- c) Keep climbing and not turn back.
- d) Ask for help from others.

Answer: c) Keep climbing and not turn back.

4. Which line from the poem shows that the mother wants her son to persevere?

- a) "I've been a-climbin' on"
- b) "Don't you set down on the steps"
- c) "Life for me ain't been no crystal stair"
- d) "And sometimes goin' in the dark"

Answer: b) "Don't you set down on the steps"

Note. This figure illustrates how applying the RAFT (Role, Audience, Format, Topic) strategy improved the quality of content generated by ChatGPT.

The questions generated through this approach were relevant and effectively targeted the comprehension skills we aimed to develop, such as identifying explicit details and interpreting textual evidence. The AI-generated questions aligned precisely with our objectives by centering on key lines from the poem and prompting students to link these details to broader themes. Using RAFT allowed us to achieve these results quickly by streamlining the content creation process.

Even with RAFT, we found that follow-up prompts were often necessary to focus on specific content areas or skills, such as emphasizing figurative language or ensuring the inclusion of particular vocabulary. For example, after generating a set of basic comprehension questions, we might provide more specifics about vocabulary or skills, as seen in Figure 5. In this way, follow-up prompts ensured that AI-generated materials were not only relevant but also aligned with instructional priorities, reinforcing key concepts and scaffolding student learning effectively.

Figure 5: *Refining AI-Generated Content with Follow-Up Prompts*

Refine the questions to focus on figurative language used in the poem, specifically metaphors and symbolism.



Here are refined multiple-choice questions focusing on the figurative language, particularly metaphors and symbolism, used in "Mother to Son" by Langston Hughes:

Questions

1. What is the primary metaphor used throughout the poem "Mother to Son"?
 - a) A crystal chandelier
 - b) A staircase
 - c) A rocky mountain
 - d) A smooth road
2. In the poem, the "crystal stair" is a symbol for what?
 - a) A life of luxury and ease
 - b) A difficult journey
 - c) A steep climb
 - d) A broken relationship
3. What do the "tacks" and "splinters" in the staircase symbolize?
 - a) Beautiful decorations
 - b) The mother's childhood memories



Note. This figure shows how follow-up prompts were used alongside the RAFT strategy to fine-tune AI-generated content.

This iterative approach mirrored the dynamic nature of classroom teaching, where instruction is continually adjusted based on student needs and responses. By prompting ChatGPT to include specific vocabulary or focus on particular themes, we could ensure the AI-generated content was targeted and comprehensive, thus better supporting our students' learning progression. Engaging in this back-and-forth exchange with ChatGPT also mimicked the collaborative nature of lesson planning, where refining and revising are essential to creating high-quality educational materials. Ashley noted that ChatGPT had become a third non-human member of our PLC.

By refining our prompts with strategies like RAFT and engaging in a continuous dialogue with ChatGPT, we transformed it from a basic question generator into a customized assessment co-designer. This collaborative process allowed us to leverage the AI's capabilities to ensure the generated content was relevant and aligned with our instructional goals. Our ability to iteratively refine the prompts and revise the output ultimately enabled us to create more precise, meaningful assessments and instructional materials that directly supported our students' learning, enhancing comprehension and critical thinking skills in previously time-consuming and challenging ways.

Key Teacher Moves for Effective AI Engagement to Refine Assessment Items

1. Provide Role, Audience, Format, and Topic (RAFT) in initial prompts.
2. Name specifics for what you want.
3. Engage in iterative dialogue with GenAI.
4. Evaluate and revise output.

Using AI to Simplify Complex Texts

As we became more proficient in using ChatGPT for assessment creation, we recognized its potential to simplify complex texts and create more accessible learning materials for our students. This practice was particularly valuable when working with historical and legal texts, which often contained dense language and intricate concepts. For our students reading below grade level, accessing these materials in their original form posed a significant challenge. Simplifying these texts while maintaining the core ideas was crucial for ensuring all students could engage with grade-level content and participate in higher-order thinking. As seen in Figure 6 (next page), this simplified version retained the original text's core concepts while translating legal jargon into language accessible to students reading at a lower grade level. By making the text more approachable, we enabled our readers to build foundational knowledge that prepared them for engaging with more sophisticated ideas. Before using ChatGPT, adapting materials for our readers was a labor-intensive process that could take several hours. With ChatGPT, this process was streamlined, allowing us to reallocate time toward more dynamic and interactive instructional activities.

Figure 6: *Simplifying Complex Texts for Accessibility*

Revise this to make more accessible for 8th graders reading on a 5th grade level. Explain concepts that might be difficult: On the Fourteenth Amendment question, the Court conceded that the object of the Amendment was to enforce “political” equality of the races “before the law.” But the Court then advanced two sweeping propositions. First, it said that the Fourteenth Amendment “could not have been intended to abolish distinctions based upon color, or to enforce social, as distinguished from political, equality, or a commingling of the two races upon terms unsatisfactory to either.” Second, the Court argued that laws requiring segregation of the two races did not necessarily imply inferiority of either.



On the Fourteenth Amendment issue, the Court agreed that the purpose of the Amendment was to make sure all races were treated equally by the law. This means everyone should have the same legal rights, like voting and being treated fairly in court. However, the Court then made two broad statements.

First, it said that the Fourteenth Amendment was not meant to erase all differences based on skin color or to make people of different races mix socially if they didn't want to. In other words, the Court believed the Amendment was about legal rights, not forcing people to be friends or interact in social settings.

Second, the Court argued that laws that kept races separate, like having different schools or bathrooms for Black and white people (segregation), did not necessarily mean that one race was seen as worse or less important than the other. They believed it was possible to separate races without saying one race was better than the other.

Note. This figure illustrates how ChatGPT was used to simplify complex texts to make them more accessible for students reading below grade level.

Our dialogue with the AI iteratively refined outputs until they were optimized for our instructional purposes. For example, in the text about the Fourteenth Amendment above, we appreciated how the introductory paragraph highlighted the presence of “two broad statements,” reinforcing this structure with clear linguistic cues—“First” in one paragraph and “Second” in the next. These signals would support students' comprehension. Additionally, we valued how the revised text incorporated the word “segregation.” In our next iteration, we prompted ChatGPT to include “abolish” and “inferiority,” as these were essential terms for students to understand and use. By leveraging our professional judgment and experience, we guided ChatGPT, transforming it from a basic text simplification tool into a strategic partner in enhancing student learning.

We applied our understanding of the content and our students' specific learning needs to use ChatGPT to simplify complex texts. In our PLC meetings, we used RAFT to thoughtfully design prompts that communicated our instructional objectives and specified the format of the outputs. We evaluated the AI-generated outputs, ensuring they were simplified, educationally sound, and aligned with our goals.

Key Teacher Moves for Effective AI Engagement for Simplifying Complex Texts

1. Create prompts with careful consideration of the desired output.
2. Use iterative refinement.
3. Apply professional judgment—drawing on your expertise in pedagogy, content, and student needs—to review and adjust AI-generated output to align with your instructional goals.

Utilizing AI for Strategic Instructional Planning & Student Support

As we worked with ChatGPT, we realized that our approach was evolving from using it as a product production platform to a creative thinking tool. Initially, we viewed ChatGPT much like an assistant—generating questions and simplifying texts. However, as we became more confident, we engaged in a "conversation" with the AI, refining and adjusting its outputs much like in a dialogue with a teaching colleague. This evolution was not automatic; it required us to continuously apply our knowledge of the curriculum, effective instructional strategies, and student needs—and how to protect student confidentiality—to guide ChatGPT effectively. The AI was a powerful tool, but it only became pedagogically sound when we actively directed and refined its responses to align with our instructional goals.

Scaffolding Instruction

One of the first instructional strategies we explored with generative AI focused on enhancing students' inferencing skills. Many of our sixth-, seventh-, and eighth-grade students struggled in this critical area. While we had become proficient in using ChatGPT to generate questions for comprehension, we noticed that students often found it challenging to make accurate inferences, especially when they had varying levels of background knowledge. Inferencing requires students to read between the lines, drawing on textual evidence and their understanding to form conclusions or predictions, and we realized this was an area that needed more focused development.

During several PLC meetings, we discussed the specific knowledge and cognitive processes students needed to answer inference questions successfully. We also reflected on the steps we followed when making inferences. These conversations led us to revisit a strategy we had recently encountered in professional development: three-level comprehension guides (Vacca, Vacca, & Mraz, 2013). This strategy involves using three levels of comprehension—literal, interpretive, and applied—to scaffold students' understanding and inferencing skills.

We decided to experiment with this approach using ChatGPT. We investigated ChatGPT's familiarity with the three-level comprehension strategy, and upon receiving a clear, accurate description that aligned with our understanding, we set out to create instructional materials. We began by having ChatGPT generate three-level comprehension statements using familiar and accessible texts—like the classroom environment as the "text." For instance, a literal statement might be, "Ms. Ashley posts students' birthdays on the wall," which students could confirm with visual evidence. This would then lead to an interpretive statement such as, "Ms. Ashley values her students as individuals," based on the evidence of personalized engagement and classroom activities. The final applied statement could require students to predict how Ms. Ashley might respond if a new student joins the class mid-year, drawing on their understanding from the earlier levels.

As we worked with ChatGPT to generate three-level statements, we realized that the generated statements often did not always logically build from one level to the next. To address this, we reversed our approach, starting with the applied or inference-based statement and asking ChatGPT to work backward to create the interpretive and literal statements. This method proved more effective in ensuring that each level of comprehension logically supported the next, guiding students step-by-step from basic understanding to deeper inference.

Figure 7: Refining Three-Level Statements for Deeper Comprehension

Write level 1 and level 2 questions to help students answer this level 3 question: Based on the passage, what values or beliefs do you think are important to the narrator's family during Tet?



Sure, here are level 1 and level 2 questions that can help students answer the level 3 question about the values or beliefs important to the narrator's family during Tet:

Level 1 Question (Recall/Understanding):

- What is Tet, as mentioned in the passage?
- What actions does the family take on Tet?
- Why does the narrator's mother warn them about their behavior on Tet?

Level 2 Question (Application/Analysis):

- How do the actions described in the passage contribute to the atmosphere of Tet?
- What do you think is the significance of the narrator's mother warning them about their behavior on Tet?
- How do the actions and beliefs mentioned in the passage reflect the importance of tradition and family unity during Tet?

These level 1 and level 2 questions provide students with the foundational understanding and analytical skills necessary to infer the values or beliefs of the narrator's family during Tet, which is the focus of the level 3 question.



Note. This figure shows how reversing the process of generating three-level statements with ChatGPT improved logical progression from one comprehension level to the next. The AI generated a more coherent set of statements by starting with an inference-based statement and working backward to create interpretive and literal statements.

As we refined our approach, we experimented with mixing statements students could affirm or refute based on the text with questions they had to answer, all at the same comprehension level. For example, we asked students to affirm or refute the statement: “Puck, rather than fate, is responsible for the chaos in the play.” We then followed up with ChatGPT generated: “If Puck causes the chaos, what does this say about Shakespeare’s view on fate versus choice? If he doesn’t, what else creates the disorder, and why does it matter?” This combination helped students build a more layered understanding of the text, applying knowledge to more complex inference questions. The AI’s ability to quickly generate multiple practice opportunities enabled us to provide students with exercises that gradually increased in complexity, reinforcing their skills over time.

This approach allowed students to engage in regular practice with challenging texts and to become familiar with the structure of the three-level comprehension guides. Students began to recognize these patterns, often referring to them as “3CG sets,” and understood that each question was a stepping stone toward answering the final, most challenging question. They developed a clearer appreciation for returning to the text—a practice embedded in the level one questions—as a foundational strategy for answering higher-level inference and application questions. Over time, we observed that students improved in answering literal-level questions correctly, which provided a strong baseline for tackling more complex questions. Their accuracy increased across all three question levels, and they demonstrated greater confidence and willingness to engage in classroom discussions around level three questions, using textual evidence to support their reasoning.

By leveraging generative AI, we created targeted, scaffolded practice opportunities more efficiently than ever. Through an iterative process of refining and adjusting the AI’s outputs, we developed a series of comprehension exercises that built on students’ existing knowledge, guiding them from basic understanding to more complex inferencing. This approach enhanced our students’ skills and fostered a deeper understanding of navigating texts systematically, drawing on evidence and reasoning to make informed inferences. Using ChatGPT for this type of lesson planning, we could tailor these activities to support our students’ growth.

Supporting Meta-cognitive Skills & Student Empowerment

As we continued experimenting with ChatGPT, we recognized its value not just as a content generator but as a tool to support the development of meta-cognitive skills. Through our classroom observations, we noticed that while our students were adept at comprehending concrete information and literal meanings and were finding success with the three level-level questions, they often faced challenges when inferring deeper motivations from more less common cues. Our students struggled with higher-order thinking skills, like those described by Theory of Mind (ToM)—the ability to understand that others have different thoughts, beliefs, and intentions (Dore et al., 2018). This gap was particularly evident during discussions that required our students to interpret character motives, anticipate actions, or understand irony and implied meanings. For instance, when asked why a character might behave in a certain way, students frequently provided surface-level responses that overlooked the character's internal states or broader contextual factors. For example, when discussing Egeus's motives in *A Midsummer Night's Dream*, a common response was simply, "Egeus is mad at Hermia because she won't listen to him." While technically correct, this response failed to consider his belief in patriarchal authority, his need for control, or his personal stake in Hermia's marriage to Demetrius. We realized that for our students to engage with and comprehend narrative texts fully, they needed to develop stronger ToM and related meta-cognitive skills.

To address this challenge, we used ChatGPT to create materials to enhance our students' ToM and critical thinking abilities. One effective approach involved introducing students to Happe's (1994) "Strange Stories," which are crafted to test and develop ToM by presenting scenarios that require students to predict or explain a character's behavior based on their mental state. Recognizing the importance of varied and repeated practice to strengthen these skills, we turned to ChatGPT to generate additional stories similar in style and structure to Happe's.

Our process began by engaging ChatGPT in a dialogue about ToM, prompting it to explain the concept and provide relevant examples. This exchange helped confirm our understanding and allowed us to ensure that ChatGPT's responses aligned with current educational research. Although the version of ChatGPT we used at the time did not have direct access to several specific articles we had studied (e.g., Dore et al., 2018), its responses were generally consistent with the concepts presented in the literature, reinforcing its utility for our purposes.

We provided ChatGPT with "Strange Stories" excerpts found on the Internet (used for educational purposes under fair use principles) to serve as examples to guide the generation of new scenarios. Using examples rather than abstract descriptions led to more relevant and accurate outputs. ChatGPT's ability to quickly generate multiple new "strange stories" allowed us to offer our students a broader range of materials as well as personalized stories to practice their ToM skills.

In addition to building ToM, we aimed to strengthen our students' questioning skills, a critical component of reading comprehension and meta-cognition. To facilitate this, we used ChatGPT to generate a series of questions related to each story, as seen in Figure 8. Rather than having students simply answer these questions, we focused on fostering meta-cognitive reflection by asking them to discuss which questions would be most helpful in understanding the stories and why. This approach encouraged students to think critically about the purpose and value of different types of questions.

Figure 8: *Enhancing Questioning Skills & Meta-Cognitive Reflection*

Write three questions students could ask that would help them understand this story.



Here are three questions that students could ask to help them understand the "Misunderstanding" story:

1. Why does Mrs. Peabody react with fear when she sees the man approaching her?

This question encourages students to consider Mrs. Peabody's perspective and understand her fear of being attacked, which is based on her previous experiences and anxieties.

2. What might the man be thinking or feeling when Mrs. Peabody offers him her purse?

This question prompts students to think about the man's perspective and his possible confusion or surprise at Mrs. Peabody's unexpected reaction.

3. How could the situation have been different if Mrs. Peabody had not been so afraid?

This question encourages students to think about how Mrs. Peabody's assumptions and feelings influenced the outcome of the interaction and consider alternative scenarios.

Note. This figure demonstrates the use of ChatGPT to generate questions aimed at strengthening students' questioning skills and meta-cognitive reflection.

We utilized these ChatGPT-generated questions to engage students in discussions about which questions would deepen their understanding of the characters and which might be less useful. Interestingly, students exceeded our expectations by thoughtfully debating the relevance of certain details, such as Mrs. Peabody's age, in understanding her reaction to a man's approach. They reasoned that her response could vary significantly depending on whether she was very old or young, demonstrating a deeper level of inference and engagement than we had initially anticipated. Building on this foundation, we then had students generate their own questions about the stories. To support varying needs, some students received question stems as scaffolds, while others generated questions independently.

By emphasizing higher-order thinking, we continued to move beyond using ChatGPT as an efficiency tool. Instead, we engaged in a dynamic, iterative process with the AI, continuously refining its outputs to design richer, more meaningful learning experiences. This collaborative approach allowed us to tailor materials specifically targeting metacognitive skills. Through this dialogue with ChatGPT, we adapted our instructional strategies to better support our students' unique needs.

Identifying & Anticipating Learning Challenges

We discovered that ChatGPT could be a valuable tool for proactively identifying and anticipating potential learning challenges, allowing us to develop targeted teaching strategies. Understanding where students might struggle was crucial in crafting lessons that effectively supported their learning needs. For instance, when we introduced a complex commentary on the Fourteenth Amendment, we anticipated that the dense language and intricate historical context could create significant barriers to comprehension for our students.

To address this, we used ChatGPT to help pinpoint specific areas where students were likely to face difficulties. See Figure 9. The AI provided insights into potential vocabulary hurdles and conceptual misunderstandings that could arise from the text. With the feedback from

ChatGPT, we engaged in discussions during our PLC meetings to determine the most impactful starting points for instruction. We decided to focus on building students' vocabulary as a foundational step, recognizing that a strong grasp of key terms was essential for understanding the broader context.

Figure 9: *Using AI to Identify Instructional Focus Areas*

Consider this commentary on the 14th Amendment: On the Fourteenth Amendment question, the Court conceded that the object of the Amendment was to enforce "political" equality of the races "before the law." But the Court then advanced two sweeping propositions. First, it said that the Fourteenth Amendment "could not have been intended to abolish distinctions based upon color, or to enforce social, as distinguished from political, equality, or a commingling of the two races upon terms unsatisfactory to either." Second, the Court argued that laws requiring segregation of the two races did not necessarily imply inferiority of either. What background knowledge would an 8th grader need to understand this paragraph? How might a student struggle to make meaning?



To understand this paragraph about the 14th Amendment, an 8th grader would need several pieces of background knowledge:

Background Knowledge Needed:

1. **The Fourteenth Amendment:** Understanding that the 14th Amendment, ratified in 1868, was one of the Reconstruction Amendments adopted after the Civil War. It was designed to grant citizenship to all persons born or naturalized in the United States, including former enslaved people, and to provide all citizens with "equal protection under the laws."
2. **Concepts of "Political" vs. "Social" Equality:** Knowledge that "political equality" means



equal treatment under the law, such as the right to vote or access public services. "Social equality," on the other hand, refers to social integration and equal treatment in everyday social life, which includes things like marriage, education, and public accommodations.

3. **Historical Context of Segregation:** An understanding of racial segregation laws (Jim Crow laws) that mandated the separation of races in public and private spaces in the United States, primarily in the South, after the Civil War and into the 20th century.
4. **Legal Reasoning and Court Decisions:** Basic knowledge of how the U.S. Supreme Court interprets laws and constitutional amendments, including landmark cases like *Plessy v. Ferguson* (1896), which upheld the constitutionality of racial segregation under the doctrine of "separate but equal."
5. **Terms Like "Inferiority" and "Distinctions Based on Color":** Familiarity with the notion that segregation was often justified with arguments that it did not imply one race was inferior to another, despite the social implications and lived realities suggesting otherwise.

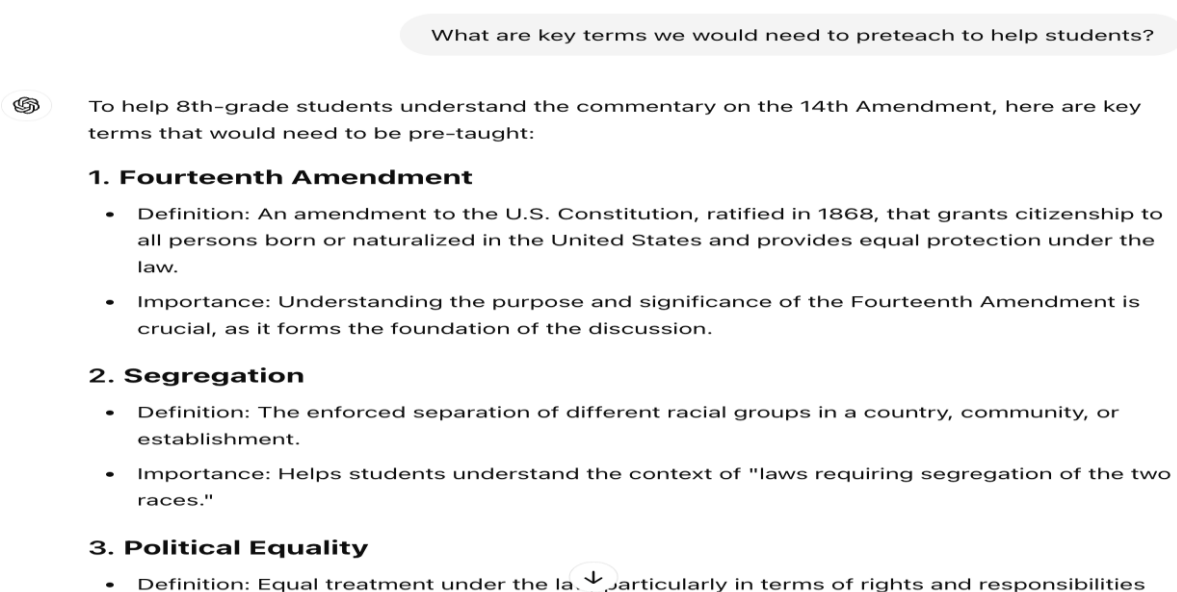
Potential Struggles for an 8th Grader:

1. **Complex Legal and Philosophical Language:** The paragraph uses legal and philosophical terms like "political equality," "social equality," "commingling," and "inferiority," which may be abstract or unfamiliar to an 8th grader. These terms require a nuanced understanding of both law and social concepts.
2. **Understanding Court Reasoning:** The reasoning of the Court, especially the idea that laws requiring segregation do not imply inferiority, may be confusing. Students might struggle with the contradiction between the claim of legal equality and the reality of social segregation and

Note. This figure illustrates how ChatGPT was used to identify specific areas where students might face difficulties, such as challenging vocabulary and potential conceptual misunderstandings.

As we refined our instructional approach, we continued to use ChatGPT to generate vocabulary lists and definitions tailored to the text's context. As illustrated in Figure 10, using AI to generate and differentiate vocabulary lists helped scaffold learning.

Figure 10: Generating & Differentiating Vocabulary Lists Using AI



such as voting, access to public services, and legal protection.

- Importance: Clarifies the Court's use of "political equality" as distinct from "social equality."

4. Social Equality

- Definition: Equal status and integration of all people in social settings, which includes rights related to social interactions, access to public places, and intermingling in society.
- Importance: Understanding this term helps students grasp the distinction made by the Court between legal rights and everyday social interactions.

5. Inferiority

- Definition: The condition of being lower in status or quality than another or others.
- Importance: Helps students understand the Court's argument that segregation laws did not imply one race was inferior to the other.

6. Distinctions Based on Color

- Definition: Legal or social separations or categorizations made on the basis of skin color or race.
- Importance: Key to understanding the Court's stance that the Fourteenth Amendment was not intended to abolish racial distinctions.

7. Commingling



Note. This figure shows how ChatGPT was utilized to generate vocabulary lists and definitions tailored to the context of the text.

This use of ChatGPT allowed us to foresee learning obstacles that might have otherwise required extensive time, observation, and trial-and-error to identify in the classroom. By anticipating potential stumbling blocks well in advance, we could design lessons that directly addressed these gaps. For example, after reviewing the AI-generated list of vocabulary terms, we selected the most relevant and impactful words—such as segregation, equality, political, social, inferiority, and commingling—and identified terms that might pose comprehension challenges. Instead of waiting to see where students might struggle, we preemptively asked ChatGPT to generate student-friendly explanations and examples, particularly for more abstract words like inferiority and commingling. AI-generated analogies, such as comparing commingling to mixing different colored beads and explaining inferiority through self-esteem in competitive sports, helped us prepare targeted explanations and activities.

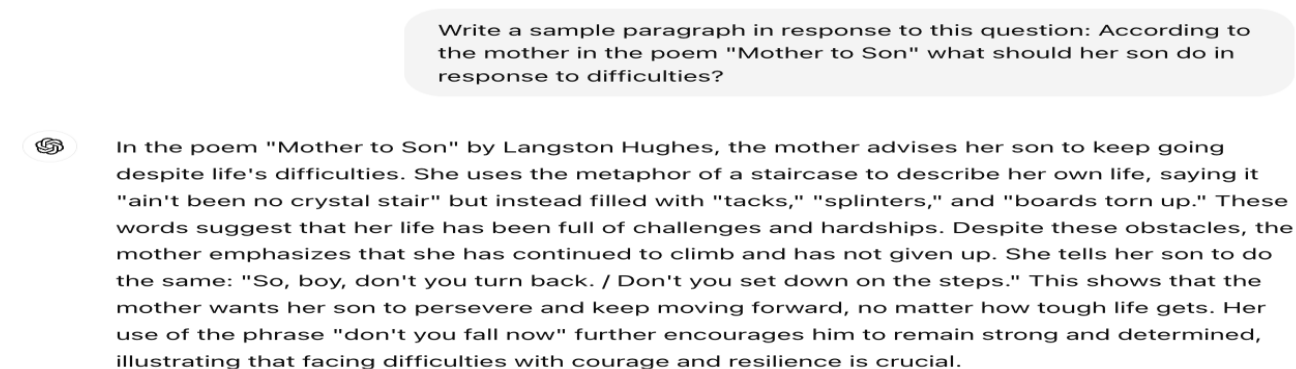
While ChatGPT provided numerous options and suggestions, our role was to critically evaluate each one, considering its relevance and applicability to our specific learning context and student needs. When the AI initially suggested a traditional word-definition matching exercise, we knew this would not effectively engage our students. Instead, we used this output as a starting point and asked ChatGPT to create a more interactive group game to reinforce word meanings. The AI-generated vocabulary charades activity, where students acted out words while peers guessed, would engage our students during class. We also prompted ChatGPT to generate sentence stems incorporating key vocabulary, ensuring that all students, including those needing more support, could use the words in context. This interplay between AI assistance and teacher expertise enabled us to create more responsive and targeted lesson plans.

Generating Exemplars to Model Skills

Creating exemplars was another area where we effectively utilized generative AI to enhance our instructional practices. Exemplars—high-quality examples demonstrating key skills—are invaluable for helping students understand expectations and develop their abilities. While we saved as many exemplars from previous students as possible, each class's unique and individualized needs meant we frequently needed to create new examples tailored to specific learning goals. This process was time-consuming, especially when catering to the diverse skill levels within a single classroom.

Recognizing the potential of ChatGPT to assist in this area, we began leveraging the AI to generate exemplars that modeled critical skills, such as citing textual evidence to support a claim. For instance, when teaching our students to analyze the poem "Mother to Son," we used ChatGPT to create multiple sample exemplar paragraphs demonstrating how to integrate and effectively explain textual evidence, as shown in Figure 12.

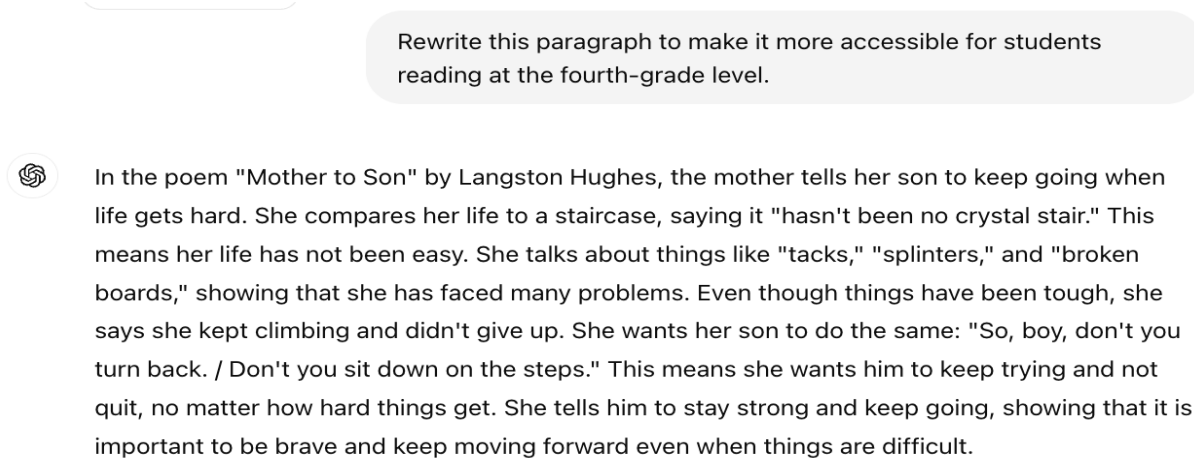
Figure 11: *Using AI to Generate Exemplars for Critical Skill Development*



Note. This figure shows how ChatGPT was utilized to create exemplar paragraphs that modeled critical skills, such as citing textual evidence to support a claim.

While this exemplar was useful, we realized that some students, particularly those reading at lower levels, might find the language inaccessible. To address this, we iteratively refined the AI-generated exemplars, prompting ChatGPT to simplify the language while preserving the key ideas and analytical rigor. See Figure 13. This process involved providing feedback to the AI and revising its outputs. This approach allowed us to produce multiple versions of each exemplar, tailored to different reading levels. By scaffolding the exemplars in this way, we ensured that all students could engage with the material meaningfully, reinforcing their understanding of constructing a well-supported paragraph.

Figure 12: *Refining AI-Generated Exemplars for Accessibility & Skill Development*



Note. This figure illustrates the process of refining AI-generated exemplars to make them more accessible to students reading at different levels.

Our ongoing interaction with ChatGPT allowed us to refine the AI's outputs continuously. We provided specific prompts and feedback, asking the AI to emphasize particular skills or aspects of analytical writing, such as the clarity of thesis statements or the depth of evidence analysis. To do this effectively, we had to rely on our expertise in writing instruction—understanding not just what makes strong analytical writing but also how to scaffold it for different levels of learners. This iterative process ensured that the exemplars modeled high-quality writing and aligned closely with our instructional goals and learning objectives. The ability to quickly generate and adapt these exemplars freed up valuable time, again enabling us to focus more on how to actively engage students with the content and guide them through the process of analyzing and interpreting texts.

Designing Multi-sensory Activities to Enhance Engagement

As our collaboration with ChatGPT deepened, we saw its potential for designing multisensory learning experiences. Multisensory activities are exceptionally vital for students in special education, as they help reinforce learning by engaging multiple senses, which enhances memory retention and understanding (Tomlinson, 2014). Despite recognizing the importance of these activities, we often struggled to find the time and resources to develop them fully. ChatGPT offered support, allowing us to explore creative and engaging ways to cultivate our students' learning through a variety of sensory modalities.

Initially, we used ChatGPT to brainstorm general ideas for multisensory activities. The AI provided a range of ideas, but some were too complex or required resources beyond our reach. However, as we iteratively refined our prompts and provided more specific information about our classroom context, ChatGPT's outputs became more targeted and practical, allowing us to develop high-quality multisensory activities.

One successful example of this collaborative design process was our work with the book *Inside Out and Back Again* by Thanhha Lai. We had identified the papaya as a significant symbol throughout the text, recognizing that it represented more than just a fruit—it was a powerful symbol of home, memory, and change for the protagonist, Ha. Given that our students had little familiarity with papayas, their growth stages, or their cultural significance and knowing our students

themselves, we knew that a lecture would not suffice to convey the depth of this symbolism. We needed to create a multisensory experience to engage students and help them build a richer understanding.

Ashley had already planned to bring fresh papayas and dried papaya slices into the classroom, allowing students to touch, smell, and taste the fruit. This direct engagement provided a tangible connection to the text, making the papaya more than just an abstract concept. Alongside this, ChatGPT recommended using images and videos showing the stages of papaya growth, from seedling to ripe fruit. These visual aids helped students visualize the fruit's development, connecting it to Ha's memories and experiences as described in the poem "Papaya Tree."

To support students' understanding of symbolism, we collaborated with ChatGPT to develop a multi-sensory activity centered around the papaya tree. ChatGPT suggested an art-based mapping activity, prompting students to create visual representations of the tree and incorporate their interpretations of its meaning as presented in the text. This activity engaged our learners, allowing them to express their understanding creatively while reinforcing their grasp of literary symbolism. We further refined this idea by leveraging ChatGPT's capabilities to generate discussion prompts. These prompts encouraged students to explore how the papaya might symbolize different concepts, such as hope, loss, or resilience, in various parts of the story.

Figure 13: Enhancing Symbolism Understanding Through Art & Role-Playing Activities

You are a special education teacher working with 7th graders who read on the 4th grade reading level. They are reading *Inside Out and Back Again*. What are two multisensory activities that could help them understand the poem "Papaya Tree." I also want to prepare them to understand the ongoing symbolism of the fruit throughout the book.



Here are two multisensory activities to help 7th graders who read at a 4th-grade level understand the poem "Papaya Tree" from *Inside Out and Back Again* and prepare them to recognize the ongoing symbolism of the fruit throughout the book:

1. Papaya Tree Sensory Exploration and Symbolism Mapping

Objective: Engage students with the sensory details in the poem "Papaya Tree" to deepen their understanding of the imagery and symbolism of the papaya fruit throughout the book.

2. Papaya Tree Art and Storyboard Creation

Objective: Help students visualize and articulate their understanding of the symbolism of the papaya tree and its significance in *Inside Out and Back Again*.

Activity Steps:

1. Visual Representation of the Papaya Tree:

- Start the lesson by reading the poem "Papaya Tree" together. Discuss the imagery and emotions conveyed in the poem.
- Give students large sheets of paper or poster boards and art supplies. Ask them to create a visual representation of the papaya tree as described in the poem. They can draw, color, or use collage materials to depict the tree, the fruit, and any other elements mentioned.
- Encourage students to incorporate words or phrases from the poem into their artwork that capture the feelings and themes associated with the papaya tree (e.g., "home," "harvest," "home," "home")

Note. This figure shows excerpts from a ChatGPT conversation, demonstrating how it was used to suggest multi-sensory activities that deepened students' understanding of the symbolism of the papaya in a text.

Building on this foundation, ChatGPT helped us develop an interactive storytelling and role-playing exercise. Students worked on storyboarding key scenes from the book where the papaya is mentioned, discussing its evolving symbolism. The AI provided suggestions for role-playing scenarios where students could act out these scenes, allowing them to embody the characters' emotions and experiences.

By collaborating with ChatGPT to design multisensory activities like these, we engaged our students far more deeply than through traditional methods. The sensory experiences helped students build background knowledge and facilitated a deeper comprehension. After the papaya activities, students were able to describe the fruit's significance in relation to Ha's personal journey, a notable improvement from earlier discussions where responses were often vague or limited to surface-level observations. When we revisited the poem "Papaya Tree" later in the unit, students demonstrated greater retention and textual recall.

This experience exemplifies a larger shift in our instructional approach, demonstrating how ChatGPT allowed us to expand our lesson planning in new, creative directions. The AI's ability to quickly generate diverse, multisensory activities tailored to our specific classroom context enabled us to plan more effectively and engage our students more fully.

Key Teacher Moves for Effective AI Engagement

1. Engage in iterative dialogue
2. Provide clear and specific prompts.
3. Collaborate with AI as a co-designer.
4. Evaluate and revise AI outputs.

Conclusion

This paper has reflected on our journey integrating generative AI, specifically ChatGPT, into instructional planning within a special education context. Throughout our experiences, we

identified several critical teacher moves essential for effectively leveraging AI tools, such as engaging in iterative dialogue, providing clear and specific prompts, and rigorously evaluating and refining AI outputs. These strategies are central to maximizing the pedagogical value of AI and align closely with the broader framework of Technological Pedagogical Content Knowledge (TPACK) (Mishra & Koehler, 2006). The TPACK framework emphasizes that effective engagement with generative AI depends on a solid foundation of teaching expertise across content knowledge, pedagogical knowledge, and technological knowledge. Key Teacher Moves aligned with the TPACK Framework are listed in Table 1.

Table 1: *Key Teacher Moves within the TPACK Framework*

Key Teacher Moves within the TPACK Framework:	Description
Engage in Iterative Dialogue	Continuously interact with AI, refining prompts and responses to ensure outputs align with instructional goals. This process mirrors refining lesson plans in dialogue with a colleague, blending technological tools with pedagogical strategies to enhance learning.
Provide Clear and Specific Prompts	Develop prompts that clearly communicate instructional objectives to enhance the relevance and quality of AI-generated content. Understanding both the technological tool (AI) and the content area is essential to generate appropriate and accurate materials.
Define the Output Structure	Clearly specify the desired format and content structure of AI outputs to ensure alignment with educational standards and instructional goals. This requires knowledge of both the technological capabilities and the specific content requirements.
Ask for More than Needed	Request more AI-generated examples than necessary to provide a broader selection and reduce the need for multiple refinement rounds. This strategy leverages an understanding of AI's capabilities to maximize its utility.
Evaluate and Revise AI Outputs	Critically assess AI outputs for quality and appropriateness, making adjustments to align with learning objectives and pedagogical soundness. This move requires an integrated understanding of how technology can support pedagogical practices and content delivery.

Within and beyond these Key Teacher Moves, our experience demonstrates that effective integration of AI in education is not merely about using technology but involves a nuanced application of teacher knowledge to guide AI's use. Understanding the curriculum and standards was crucial for prompting the AI accurately and ensuring that the generated materials aligned with our

instructional goals. Similarly, knowing our students' diverse needs allowed us to tailor AI outputs to make them accessible and meaningful.

Our reflection underscores the importance of recognizing AI's limitations, including potential biases in outputs, and the critical role of teacher expertise in mitigating these challenges. It also highlights the necessity for ongoing professional development to equip educators with the skills to effectively integrate AI tools into their teaching practice, ensuring they are used appropriately, ethically, and effectively to support diverse learners.

Ultimately, our exploration of generative AI in the classroom reveals a significant shift—from using AI as a tool for efficiency to engaging it as a non-human co-designer in creating richer, more meaningful learning experiences. This journey has shown us that AI's true potential in education is unlocked when teachers apply their expertise to thoughtfully guide and refine its outputs.

References

- Choi, G. W., Kim, S. H., Lee, D., & Moon, J. (2024). Utilizing generative AI for instructional design: Exploring strengths, weaknesses, opportunities, and threats. *TechTrends*, 68(832–844). <https://doi.org/10.1007/s11528-024-00967-w>
- Dickey, E., Bejarano, A., & Garg, C. (2024). AI-Lab: A framework for introducing generative artificial intelligence tools in computer programming courses. *SN Computer Science*, 5, 720. <https://doi.org/10.1007/s42979-024-03074-y>
- Dore, R. A., Amend, S. J., Golinkoff, R. M., & Hirsh-Pasek, K. (2018). Theory of mind: A hidden factor in reading comprehension? *Educational Psychology Review*, 30(3), 1067–1089. <https://doi.org/10.1007/s10648-018-9443-9>
- DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. National Educational Service.
- Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children. *Journal of Autism and Developmental Disorders*, 24(2), 129-154. <https://doi.org/10.1007/BF02172093>
- Individuals with Disabilities Education Act, 34 C.F.R. § 300.39(b)(3) (2006). *Code of Federal Regulations*. Retrieved from <https://www.ecfr.gov/current/title-34/subtitle-B/chapter-III/part-300>
- Khazanchi, R., & Khazanchi, P. (2021). Artificial intelligence in education: A closer look into intelligent tutoring systems. In *Handbook of research on critical issues in special education for school rehabilitation practices* (pp. 256-277). IGI Global.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for integrating technology in teacher knowledge. *Teachers College Record*, 108(6), 1017-1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). ASCD.
- Vacca, R. T., Vacca, J. L., & Mraz, M. E. (2013). *Content area reading: Literacy and learning across the curriculum* (11th ed.). Pearson.

The Stakes Are High. Are the Benefits Bountiful?: HBCU Students, AI, & the Power of Composition

Adrienne Carthon¹

Abstract

An outgrowth of work submitted for Howard University's Generative Artificial Intelligence (AI) Writing Faculty Task Force, this essay examines what is at stake for students of color with the use of AI as well as the potential opportunities it presents. The research is tailored to explore how the mission of Historically Black Colleges and Universities aligns or diverges from engagement with this latest technology and its implications for the larger community. Building on existing research, this essay documents some of the conversations occurring at the margins of a much larger discourse and broadens the understanding of the intersections of technology, race, and class. While there are statements on the inherent bias of AI as it relates to communities of color and African American speech patterns and topics, this study extends to what it means for the student in the composition classroom and beyond and how HBCUs are uniquely suited to address issues of bias. I propose that HBCUs create a Generative AI product that builds on existing technology that is fueled by culturally diverse data sets that account for a truly diasporic expression of Blackness that is rooted in the lived experiences of its participants.

Keywords: *Generative AI, HBCUs, composition classroom, ChatGPT, algorithmic bias, Howard University*

This musical rhetoric is my literary inheritance. My job is to assure it never dies, to make sure my characters sound like my people in all their beauty and complexity. Yet having been taught the English language by teachers who don't know my linguistic legacy, I've had to break rules and resist conventions that promise to drown out voices I know so well. Well, I pray I have done it well.

-Daniel Black, "Rhythm in Writing," *How We Do It*

1. I have no known conflict of interest to disclose. I used ChatGPT to translate a Standard American English passage. The tool was used to provide examples of its ability to adapt to varying speech patterns. The output from the tool was not modified.

Man Versus Machine

Racism exists in the world of Artificial Intelligence (AI) because it exists in the world that creates it, duplicating and reflecting the same biases. This is a now commonly acknowledged and foregone idea. Generative AI tools such as those that assist with composition are at their weakest when met with requests related to African American culture and language, so how is it useful for composition students at Historically Black Colleges and Universities (HBCUs)? The technology is not going anywhere, but how is it useful to these students?

In the Fall 2023 semester, the director of First Year Writing at Howard University, Dr. David Green, organized a task force committee to study Artificial Intelligence (AI) technology and what its implications might be for the first year writing student population. For an academic year, the other members of the task force and I researched, met regularly and compiled all of the data that we could find on AI and composition in general and AI and HBCUs specifically. On composition, we found a quickly budding field.² Much of what we found initially focused on the use of AI as a tool, and indeed, the general tone of meetings, at first, seemed to favor how the technology could be useful in assisting composition instructors and improving student writing. However, our intended focus had to shift back to how use of this technology would positively or negatively affect members of underserved communities that mirrored the population of our institution, which led to my authoring of what I called “The Dissent” for our departmental training module on AI.

By our return from the Winter break, most of us were more aligned with ideas of dissent as we delved into discussions of plagiarism and academic integrity, not that the concerns were that AI constituted such (although it arguably does) but that there had been a shift in the national conversation. During the break, the ousting of Claudine Gay, Harvard University’s first Black president in its history, over plagiarism accusations consumed the news and raised brows in academic circles (Lawrence, A., 2024). If someone as accomplished as Gay could have her credentials questioned when AI was used to discover said instances of plagiarism that could otherwise be corrected and cited as unintentional errors, what would the backlash be for someone with considerably less training and status? (To my knowledge, Gay did not use AI; the example, however, functions to show the racialized and gendered stakes of assumptions of one not *doing the work*.) When Neri Oxman—who is not African American and is the wife of Bill Ackman, the man who led the charge against Gay, was found to have instances of plagiarism in her work, she was allowed to “apologize,” not lose her job, status, or fall from grace (Lawrence, A., 2024). The statement on a larger scale is that the rules of engagement do not apply equally across the board. Gay resigned her position over accusations that the standard response to would have been an explanation, apology, and

2. The committee consisted of Dr. Green’s assistant, Cyrus Hunter, Sean Pears, Shantay Robinson, Travis Sharp, Erin Torbett, Susan Mockler, Robin Sulkosky, Bruce Miller, and myself. At the end of the academic year, Monique Akassi, who would later replace Green as director, joined the task force and organized the inaugural conference on AI for the English Department. The link to the tool that the task force developed is <https://howardenglishai.wixsite.com/generativeai>.

correction. Had “true” plagiarism been the case, and had she truly been its worst offender?³ If one could fall from ranks so high, what was at stake for our undergraduates that instructors across the nation were poised to equip with AI literacy? In other words, would a reliance on AI cause more harm than good for our students? So, if she, at her level of experience and expertise, is not exempt from extreme scrutiny, Black students who are newly graduated from college may not stand a chance as the country was also undergoing a pervasive conservative attack on diversity, equity, and inclusion measures that questioned the legitimacy of persons of color hires and their abilities. As founded in racism as such attacks might be, they were constituting real-world consequences for people in the workplace, and if the accusations could reach a Black woman who had ascended to the highest ranks at Harvard, what would that mean for a new graduate from a leading HBCU like Howard?

An earnest literature review reveals that there is yet a dearth of peer-reviewed or published literature about artificial intelligence (AI) and the effects of its use on students of color, African Americans in particular, in composition. While there are statements on the inherent bias of AI as it relates to communities of color and African American speech patterns and topics, so far, that research does not extend to what it means for the student in the composition classroom. In *AI and Writing*, Sidney I. Dobrin, a leading rhetoric and composition scholar, posits Generative AI as a potential collaborator whose use should be acknowledged out of concern for plagiarism and a necessary reconceptualization of what that means in the AI age (Dobrin, 2023). Dobrin’s text functions as more of an introductory text to AI and writing; however, in its brief section on the challenge of bias in AI use, the bias is not figured as significant enough of a challenge to constitute the exclusion of Generative AI as a tool in composition for those who would be affected by the bias (103-110). Indeed, that’s not the purpose of the book; the book serves to provide an overview and inform the reader of the challenges that do exist with an urge that readers will use AI responsibly and ethically.

In *Teaching With AI: A Practical Guide to a New Era of Human Learning*, Jose’ Antonio Bowen and C. Edward Watson deliver what they promise is a very practical guide to understanding and using AI as an instructor who wants to equip students with a new tool for growth in academia and the workplace. As the Apostle Paul notes in the biblical text of a potential Christian convert, “*I was almost persuaded.*”⁴ In fact, the text really is useful for understanding the technology, and it is indeed the technology, and not the text, that fails the “bias test.” As the authors discuss how a liberal arts education will become more important to the necessity to frame questions for Generative AI in a way that elicits the desired responses, they cite how the creativity of a particular prompt is what results in a “creative” response by the GPT Large Language Model (LLM):

3. In the Guardian article, a political science professor debunked the notion that what was found in Gay’s work constituted academic plagiarism. Also, rhetoric and composition scholar Sidney Dobrin identifies five types of plagiarism commonly defined by academic institutions: direct, duplicate, patchwork/mosaic, paraphrasing, and accidental(34; 2023). Some of the accusations against Gay would fit into one of the more benign categorizations such as accidental.

4. In Acts 26:28, the Apostle Paul so powerfully preaches to convert King Agrippa to Christianity that Agrippa famously quips that he was almost persuaded. Similarly, I was almost sold on the virtues of AI by *Teaching With AI* until I began to test the sample prompts in ways that reflect the bias of the LLMs.

PROMPT

Write a 200-word process for removing a peanut butter sandwich from a toaster in the style of the King James Bible.

RESPONSE

Abridged from GPT 3.5 (the earlier less creative version) on July 15, 2023:

And lo, it came to pass that a peanut butter sandwich was lodged within the depths of the toaster. The people were dismayed, for they desired to partake of its savory goodness. But the sandwich was trapped, its fate intertwined with the metallic coils of the machine.

First, they did unplug the toaster from the wall, for safety was paramount... (45).

Of the LLMs, ChatGPT has emerged as one of choice for many students and professionals alike, and I was so thoroughly amused by this example that I decided to submit it to ChatGPT with some cultural variants to see the result on December 1, 2024.⁵

PROMPT

Write a 200-word process for removing a peanut butter sandwich from a toaster in the style of a person.

RESPONSE

ChatGPT

Alright, so here's how you're gonna get that peanut butter sandwich out of the toaster. First, make sure the toaster's off—safety first, right? You don't want any accidental sparks flying while you're in the thick of it.

Now, the tricky part: the sandwich is likely stuck in there pretty well because, well, peanut butter tends to get a little... gooey. You'll want to give the toaster a little shake, just a tiny nudge to see if it'll loosen up. Don't go wild, though—you don't want to send crumbs flying or make a mess.

If it's still stuck, grab a fork or a pair of tongs—nothing sharp or metal that could short-circuit the toaster. Gently, and I mean *gently*, work the sandwich free, nudging the edges where it's wedged. If it's a bit stubborn, you might need to give it a little patience.

Once it's loose, slide that peanut-buttered treasure right out of the toaster. Don't forget to keep your hands away from the heating coils—you don't want to risk a burn for a snack.

5. The first responses that I got when running the prompt variations in previous tries were lost in a document that wasn't saved, but those iterations were very similar to the ones that appear here now.

And just like that, your sandwich is free! Now, enjoy that peanut buttery goodness—just be sure to let it cool a little before diving in. (OpenAI, 2024)

PROMPT

Write a 200-word process for removing a peanut butter sandwich from a toaster in the style of a Black person.

RESPONSE

ChatGPT

Aight, so here's how you're gonna get that peanut butter sandwich out the toaster without making a mess, or worse, burnin' yourself. First thing's first: turn the toaster off. Safety is key—don't be out here playing with fire.

Now, if that sandwich is stuck, you gotta have a little patience, you feel me? Don't go reaching in with your hands—nah, that's how you end up with a toaster burn. Grab a pair of tongs, or if you don't have those, a fork will do. But nothing sharp, please, we ain't tryna short-circuit the whole thing.

Next, give it a little nudge. Be gentle, though—don't go shakin' the toaster like you're tryna start a party. You just want to get it loose. If it's real stuck, slide the tongs or fork around the edges, working the sandwich free from the coils. Take your time—this ain't a race.

Once that sandwich is loose, pull it out carefully. Don't burn yourself on them coils now! Let it cool for a second before you dive in, 'cause trust me, that peanut butter is gonna be hot.

Now, take that sandwich, sit down, and enjoy the best peanut butter creation you've ever had. You earned it. (OpenAI, 2024)

PROMPT

Write a 200-word process for removing a peanut butter sandwich from a toaster in the style of a Trinidadian person.

RESPONSE

ChatGPT

Aye, listen nah, yuh wan' take out yuh peanut butter sandwich from de toaster without any problem? Well, lemme show yuh how to do it right, so yuh don't burn yuhself or mek a mess.

First thing's first, turn off de toaster. Safety always—yuh not goin' make a mistake here, right? Now, if yuh sandwich get stuck in de toaster, don't panic,

take yuh time. Don't stick yuh hand in de toaster, yuh will burn yuhself, I tell yuh! Take a fork or tongs—nothing sharp, eh? Nuttin' that go cause a short-circuit. Yuh need to be careful.

Next, start workin' around de edges of de sandwich. Give it a little nudge, nice and slow, like yuh tryin' to free up a stubborn piece of sugarcane from de stalk. Don't rush it. If yuh sandwich stubborn, just slide de fork or tongs gently, loosening it from de wires.

Once it free, slowly pull it out. But doh forget—those coils still hot like a sun on a midday! Let de sandwich cool for a bit. If yuh too eager and bite it now, yuh gonna get a real hot surprise!

Now, enjoy yuh sandwich, partner. You did good! (OpenAI, 2024)

Notice how the response rendered like a person is written in Standard American English (SAE). According to Dobrin, “[l]anguage is another major source of bias. Most AI algorithms are designed to work in only one language—most often English (though more and more GenAI platforms are now being developed in other languages). As a result, the outputs of most GenAI programs favor data conveyed in English and are reflective of Anglophone cultures” (2023). One of the things researchers have found is that prompts have to be tailored to get the best responses from AI; however, if one is not a subject area expert or smarter than the technology, one cannot catch the errors in AI for which the user will ultimately be held accountable. Therefore, students need to know enough to check any work produced by AI that they plan to submit as/with their own. Using only Anglophone examples, stereotypical assumptions are evident in the responses to the above prompts. Only a “regular person” uses SAE, and a Black person uses a contrived version of African American Vernacular English (AAVE) according to ChatGPT. A bastardized Trinidadian version of English also applies. When shown to a Trinidadian, the country of origin for the excerpt was not apparent; like many from the Caribbean, each can identify their country's dialects when heard; this model fell short. The reason is that AI can only generate from data to which it has been exposed, resulting in exclusionary bias (Dobrin, 2023). But this is not news.

In my original “Dissent,” I note the following:

In *Poetry Will Not Optimize; or, What is Literature to AI?*, Michele Elam uses the experiment of asking AI to produce a poem in the African American tradition to test the limitations of the usage of AI. (Elam, *Poetry Will Not Optimize; or, What is Literature to AI?*, 2023) Ultimately, she concludes that creativity cannot be reduced to a prompt, suggesting that there are some things for which actual humans are still better suited. While she is careful not to be dismissive of AI completely, she still argues elsewhere in “Signs Taken for Wonders: AI, Art, and the Matter of Race” that the scientists involved in the creation and improvement of AI can benefit significantly from an education that is more “deeply integrated with the humanities and arts in order to contribute to human flourishing, particularly with regard to social justice” (Elam, *Signs Taken for Wonders: AI, Art & The Matter of Race*, 2022). Since a more thorough Humanities education, including Writing and Composition, can lead to a better qualified STEM engineer, preparing Black students to lead in that category requires a deeper dive into what Humanities faculty are already positioned and trained to do: train students who can use their own minds and develop their own voices with or without the use of AI. The onus remains on Writing faculty and others in the Humanities to teach our Black students how to write as they cannot deviate from or “play”

with rules that they do not know. Ultimately, we as teachers must continue to teach and require them to learn. There is a clear path for the emergence of a Humanities Track for the STEM student, creating a more well-rounded and well-versed student of English and Literature whose aim is to enter the STEM field as a specially trained writer and creative. So, indeed, AI does open a door, perhaps not just the one of consumerism but rather one that better serves the constituents of HBCUs. (Dissent, 2023)

The Elam essay on poetry also discusses the indeterminacy of art and the ways that literature challenges art, warning that the technology is reductive in how it views language and falls short in its “understanding” and “interpretation” of the nuances of Black English. To further illustrate the point, Elam inserts the first four lines of Maya Angelou’s “Still I Rise” as a prompt and notes how the product reveals what happens when AI meets what she terms algorithmic ahistoricity. The result is a distortion in meaning and context at the heart of questions of what is “real” and by extension, what is “human.” The essay further discusses how technology regards bias (glitch) and language systems (relational code) and how wider use does not equate to greater access and a more leveled field of power (Elam, Poetry Will Not Optimize; or, What is Literature to AI?, 2023). Yet, the algorithmic ahistoricity is extremely problematic for a population whose academic topics will undoubtedly be many times culturally-derived and self-reflective. Many of our African American and otherwise Black diasporic sources and resources are not digitized in formats to “feed the machine” of AI; however, when and if that digitization does occur, it should be to the benefit and profit of the culture that produces it. So, for now, what happens when a student needs to produce an essay that draws on the African American literary tradition? The result for an African American student is like a Jessica Care Moore title: *The Alphabet Verses the Ghetto* (Moore, 2002) because *The Words Don’t Fit in My Mouth* (Moore, 1997).

What is most bothersome about AI’s inability to address non-European American centric forms of data is that for African Americans, it replicates what African American literary theorist Houston Baker calls the “economics of slavery.” In *Blues, Ideology, and Afro-American Literature*, Baker asserts that the enslaved were grounded in an understanding of their own bodies as objects of commerce and they themselves were the chief product of rules of import and export (Baker, 1984). A similar understanding on the part of questionably free Black bodies is necessary today in that African Americans drive popular language, culture, music and mores in this country in ways that are systemically cannibalized and appropriated by communities that exist outside of their culture. Following a pattern of cannibalized innovation, very rarely are they at the top of the economic food chain in profiting from their accidental and intentional productions. For example, consider the role of Black inventors who created without the foresight to patent their work: George Crum invented the potato chip in 1883 and never patented it; and Alfred Cralle invented the ice cream scoop and never profited from his idea as well (Brown, 2023). I liken these examples to the user whom without thought for self-interacts with LLMs and other AI platforms that are fed by their usage. Students who submit their writing into LLMs and continue to tweak it via the media are, in essence, training the platforms to the seeming benefit of the end-user but to the actual financial benefit of the owners who will repurpose, repackage and resale the resulting improved product to the end-users for their further consumption. The whole process is rather brilliant when one thinks of it. However, African Americans have produced more than enough unpaid labor for this country, and as we enter yet another technological frontier, enough is enough.

Variations of English

But why should any of this matter since Standard American English *is* the preferred language of the academy? That is not necessarily the case in all composition programs. One of the central tenets of our Writing, Literacy, and Discourse course for First Year Writing at Howard University is the acknowledging the value of the different variations of English in a way that does not privilege one above all others. There is a whole field of African American Studies devoted to linguistics and the understanding of Black speech, language, and ways of meaning and understanding. Our forms of verbal and written expression go back to our collision with the English language from the transatlantic slave trade. What African Americans do with language is as improvisational as jazz itself—the cadences, rhythm, sway and mood. In the opening epigraph, Daniel Black references this. The relationship with a culturally-derived form of English, constitutive of a linguistic heritage, is one that the First Year Writing program endeavors to foster. Students are encouraged to lean into culture in the quest for their authentic voices, and more times than not, that voice does not sound like a caricature version of an assumed monolithic Blackness. The voice may sing like Trinidad, dance like Jamaica, or bop like the U.S., but for our students who are of African descent, its location is always the Black diaspora. Can AI presently keep up with that? Can AI do what Black demonstrates below:

Check out this sentence.

After sunset, Willie Joe and Bessie went to the bedroom and made love.

This sentence is okay, but it doesn't carry the rhythm of Black existence. It doesn't show or celebrate the way in which Black folks had to make time for love when the entirety of their existence was subsumed in survival. But this sentence does.

After the sun went down, Willie Joe and Bessie made their way to the bedroom and did what they could do. (Black, 2023)

This brings us back to Elam's point that there are just some things that a prompt, and by extension, Generative AI cannot do. However, I surprisingly do not find it altogether not useful at all.

Because of the educational model that produces many of our First Year Writing students whose first language is English, they are typically well-versed in Standard American English. Many of them can also execute a cookie-cutter five paragraph essay quite well as well, so the challenge of growing their writing takes a different form. They study the rhetoric and form of the works of James Baldwin, Toni Morrison, Geneva Smitherman, Ta-Nehisi Coates and a diverse group of others. They are challenged with finding their own unique writing voices. In my classroom, I encourage the students to incorporate their unique experiences that may also include their native variant or English or even another language that is not English. One model that they are given is the work of Gloria Anzaldua that does not sacrifice culture for a muted, mechanical voice.⁶ Teaching students that it is not only acceptable but desirable for their work not to sound like everyone else's demands of them that they rethink everything they have thought they have known about basic composition. We, as instructors, give our students permission to "be themselves" in

6. The work I most frequently use is *How To Tame a Wild Tongue* to illustrate how Anzaldua intersperses her English with Spanish in the service of her argument. (Lawrence A. , 2024)

their writing, and that is where Generative AI fails them. Once they begin to find their culturally distinct voices, the AI generated text does not sound like them; the distinction in text is vividly noticeable. However, issues of voice notwithstanding, it also fails to understand the nuances of regionally distinct versions of African American Vernacular English and non-monolithic expressions of Black speech. Generative AI can constitute a useful tool to enhance the student writing process for African American students to the extent that the tool can be improved to be more culturally relevant or that it is used strictly for topics that render the writers race-less.⁷

Issues of Access

We can find warnings on how technology can leave behind communities of color that lack access to it, further widening the digital class gap and creating greater inequities. However, a basic assumption is that HBCU students will have access by virtue of being college students; that may or not be universally true based on an individual institution's capacity and the individual student's access to equipment as well as the internet outside of school. Furthermore, many within the African American community take it for granted that African Americans in any field have to work twice as hard or more to get half as much in the workplace as well. And despite best efforts to increase the number of African Americans in STEM fields, a concerning shortage still exists. All of the above bring us to what is at stake in the Humanities Writing classroom with an over-reliance on AI as a resource and a tool. Tyrone Taborn argues in "The AI Revolution: Will Black America Remain a Consumer Class Forever?" that "Black representation in AI-related fields is abysmally low. Less than 5% of data scientists, machine learning engineers and tech founders are Black[.]" and "Black Americans stand on the edge of becoming permanent members of a global consumer class." Noting a necessary prioritization of programs like AI boot camps and STEM-related workshops, Taborn calls for a greater push in funding for STEM education in underserved communities. For Taborn, AI represents an opportunity for Blacks in America to level the economic playing field as the technology is as wide open for exploration as the early days of the country (Taborn, 2024). Where I find Taborn's argument most useful is in its pitch for self-determination, and that is where I see a unique role for HBCUs.

Relying on AI as a staple for writing displaces, ever so slightly (and more in some cases), the human voice and capacity for creativity. The danger for Black folks, if you will, is that while a dominant group may be lauded for its use, African Americans run the risk of its usage being another commentary on their lack or preparedness, somehow unfit and inadequate for the workplace yet again. The unfairness of the weight of that being placed on budding Black scholars is unfair, but the lack of fairness does not negate its real existence or the consequences. Additionally, teaching Black students to rely on AI also sets them up to be consumers, consumers of a medium that cannot to date accurately identify or replicate their own voices. If the mission of Howard University is to prepare leaders for a global community, then our students must also be positioned to become leaders in the field of AI. Humanities is uniquely positioned to prepare them for that.

7. Since the writing of this article, Howard University has taken the lead in addressing some of the issues raised here, partnering with Google to create an African American English data set for use by researchers and other HBCUs, and former President Ben Vinson's establishment of the Artificial Intelligence Advisory Council.

There are several problems:

1. The technology is not algorithmically set to account for our voices or cultures.
2. As it stands, we would be most likely to face punitive measures or have our intellect questioned due to the unchallenged use of the current technology and its production of hallucinations, or false data.
3. In order for the technology to account for our voices and experiences, we have to feed the machine and teach it, only for the resulting data set to benefit financially the producers (not African Americans) and not the consumers (African Americans).

In fact, many of my recent composition students consider the biases implicit in the technology grounds for its limited and guarded use in their writing; some opt not to use it all while those who do use it limit its use to grammar checking SAE and idea generation⁸.

So, what's a socially conscious HBCU to do?

Note on Commercial Theatre

By Langston Hughes

You've taken my blues and gone —
You sing 'em on Broadway
And you sing 'em in Hollywood Bowl,
And you mixed 'em up with symphonies
And you fixed 'em
So they don't sound like me.
Yep, you done taken my blues and gone.

You also took my spirituals and gone.
You put me in Macbeth and Carmen Jones
And all kinds of Swing Mikados
And in everything but what's about me —
But someday somebody'll
Stand up and talk about me,
And write about me —
Black and beautiful —
And sing about me,
And put on plays about me!

8. The students' views deserve further scholarly inquiry than the discussion here. In class discussions and journal assignments, their views ranged from not accepting of the technology at all to more moderate uses of it and everything in between. What has not been expressed is full support for the productions of Generative AI to represent the students in graded or professional writing assignments where a grade or job may be on the line. Interestingly, most lean towards writing for themselves and using the tools solely to assist with what they see as minor tasks such as grammar and mechanics. I acknowledge that the particular population that I teach may be better equipped to do so than many based on the aptitude of the average student that Howard University attracts. Understandably, efforts in other academic spaces would need to focus more readily on bringing students up to a higher level of literacy proficiency and the careful use of AI tools may be able to help them with that endeavor. It would be interesting to compare views between students of color at different HBCUs and Predominantly White Institutions (PWIs).

I reckon it'll be
Me myself!

Yes, it'll be me. (Hughes, 1990)

One potential solution is to create the LLM that we need to exist; it perhaps sounds far-fetched, but is it really? As an entrepreneur, any time I have been unable to find products to meet my needs or my clients's needs, I begin the process of creating my own. When the institutional capacity and expertise exist, why should working with generative technology be any different? Engagement with the digital humanities is an expanding field in African American Studies, and with grant funding or outside tech investor funding, it is foreseeable that a research university like Howard University could leverage its relationships with tech companies, venture capitalists, invested federal agencies, and other HBCUs to create a Generative AI product that builds on existing technology that is fueled by culturally diverse data sets that account for a truly diasporic expression of Blackness that is rooted in the lived experiences of its participants. One of the largest repositories of knowledge on Black culture exists in the Moorland Spingarn Research Center. Funding the massive undertaking of digitizing such an expansive collection would engage and employ Humanities scholars for years.⁹ In fact, the creation of such an LLM would employ numerous engineers as well and create a unique opportunity for students at HBCUs to be at the forefront of driving new technology to benefit people of color from across the globe. Imagine a model that could account for diverse ways of speaking and understanding with cultural nuance.

In the previous experiment with the peanut butter prompt, I initially asked ChatGPT to also render responses like a Geechee person and like a Gullah speaking person. I accidentally deleted the responses, hence their exclusion here, but I recall that while they were an attempt at authenticity, they still fell short. I expected as much, but imagine a creative world where a student who speaks a unique variant of English may use machine learning to help understand and speak other variants of English in an online setting that is as simple to use as a smartphone app. The resulting odds for a student from an underserved population are a lot better than using current AI and LLMs as they exist, and I am reminded of Gavin P. Johnson's cautionary message in *Don't Act Like You Forgot: Approaching Another Literary "Crisis" by (Re)Considering What We Know about Teaching Writing with and through Technologies* for teachers of composition not to act like the emergence of Generative AI in composition classrooms is the gloom, doom, and end of the profession (Johnson, 2023). Johnson quotes Bill Hart-Davidson's idea that since machine learning is already here that perhaps the role of the composition instructor is to "influence both how they work and how they are incorporated into the writing practices of people and institutions" (Johnson, 2023). To that end, the technology is as useful and meaningful as our ability to control and gatekeep the means of cultural production.

Even to my own frustration, I have wondered if the plausibility of the type of Generative AI that I propose relies too much on optimism and altruism. Fortunately, encouragement came in

9. Some such as Lawrence note the shift in the workplace caused by AI technologies, so other avenues for employment will open that require the know-how to assist or guide the technology.

the form of Calvin D. Lawrence's *Hidden in White Sight: How AI Empowers and Deepens Systemic Racism* (2023). Lawrence was the keynote speaker for Howard University English Department's inaugural AI conference, and the platform merged the practical with the academic as he shared his insight as an industry insider in the development of AI technologies. In ways that some of the texts specifically on AI and writing do not, Lawrence addresses the issues of bias in tech with rigor and condemnation while offering tangible solutions for improving the technology. His position as an academic outsider ensures that his writing is not weighed down with academic inaccessibility. Lawrence sets the chapter "Our Authentic Selves" up with the example of the plot from *Guess Who's Coming to Dinner?* (1967). The idea extrapolated from the film is that the bringing together of people from different racial and class groups actually enhances understanding and drives equitable problem-solving: "Only when we appreciate our differences enough to share them, will we find the solutions that seemingly escape us on subjects of race, love, gender, and religion" (Lawrence C. D., 2023). Taking into account something that anyone from the Black community can confirm, Lawrence notes that not all Blacks think alike, but that is also beneficial: "Corporations, and even those [...] who assemble AI teams, could learn a lot from the movie related to building diverse and inclusive teams. Of course, most would say that diversity is either in race or gender. But the film proves that just having various racial groups doesn't guarantee a diversity of opinion and thought" (Lawrence C. D., 2023). In Lawrence's experience, grouping together people who think similarly even despite their other differences results in similar problem solving styles, but "a team made up of people who think in different ways would be able to solve issues faster with more creative solutions, and their answers are less likely to harm others" (Lawrence C. D., 2023). Where Lawrence goes on to discuss a technology snafu that resulted from an AI model created by Duke University researchers is where the hairs on my arms stood up. Not to diminish the resulting bias in the Pulse model's use of an inadequate data set from Flickr Face HQ, the intrigue for me was in the act of innovation itself (Lawrence C. D., 2023). And the inadequate data set that excluded women and people of color would be less of an issue if there were institutions like, say HBCUs, working fervently and responsibly to fill the gaps of diverse data sets as it relates to AI and writing that could overflow into other areas as well. In some form, the correction is already likely occurring, but HBCUs are situated to position its students to be ahead of the curve in the creation of socially responsible AI.

In another example, Lawrence discusses the dynamics of a task force in which he participated whose goal was to address a social ill following the murder of George Floyd. In the telling of his experience, Lawrence offers one of the most impactful sections of his book. He explains that there are five core principles necessary for problem solving with AI, and he explains each of them: team diversity; culture cultivation; voice amplification and collaboration; team member contributions made visible; and empowerment. Team diversity refers to having people of diverse racial, social, gender, class and disciplinary backgrounds to help create "unbiased solutions"(95). Culture cultivation means creating an environment that allows each individual to come to the table without feeling the need to hide different facets of who they are. Voice amplification and collaboration recognizes the need to allow all invested parties the opportunity to speak and be heard on equal terms; for example, "If Black voices are not heard at the design and development table, it will be impossible for those omitted voices to weigh in when data sets are suspected of bias" (Lawrence C. D., 2023). Team member contributions made visible is an acknowledgment of the need to correct the invisibility of minority contributions in the workplace; that acknowledgement will "ensure that you have active participants in each phase of the AI design, development, and deployment process" (Lawrence C. D., 2023). Finally, empowerment is about shifting the paradigm for diverse

technologists who are multidisciplinary in nature to be comfortable and feel welcome to add their voices to the creation of AI technology that “positively impacts the world” (Lawrence C. D., 2023). Lawrence is also careful to note that we are even capable of executing bias against ourselves, so the safeguard of being surrounded by diverse voices in the creation of AI protects us all. Lawrence’s ideals are grounded in humanist thought and strengthen the argument for a Humanities-STEM major that builds on Elam’s idea that “the arts are specially poised to examine issues in technological spaces (from industry to STEM education) of equity, diversity, social justice, and power more capaciously and cogently than the sometimes reductive industry-speak of inclusion, fairness, or safety (usually simply meaning minimization of harm or death—a low bar indeed)” (“Signs” Elam; 2022).

The creation of a culturally-relevant Generative AI product by an HBCU conglomerate shifts the needle from consumerism to production as a means of repairing some of the inequities and leveling the playing field within technology. The funding around such a project would have to reach a grassroots level as Taborn’s article suggests. The role of the HBCU would be central and essential to educating large numbers of people on what Taborn calls the new frontier of AI, blockchain, quantum physics, and the metaverse (Taborn, 2024). A transformative Generative AI product with “soul” that works to mitigate harm and minimize bias that is trained by data sets compiled by HBCU students and diverse community partners is within the realm of possibility. However, if producing a culturally-sensitive and culturally-relevant form of Generative AI seems too far from reach, not near enough in the future, then what is the solution for those leading the composition classrooms filled with students of color from historically underrepresented and under-considered groups? Do the work. Ground students so well in the works of their culture that they can spot factual errors for themselves. Encourage them to embrace their unique voices with pride. Promote literacy of diverse forms of English with an understanding of SAE so firm that the HBCU Humanities-STEM graduates can advise those who would program the machines or program them themselves. Lead them in discussions of ethics that will prepare them to safeguard the data they compile from weaponized abuse. Prepare them to contribute to and shape the conversations surrounding AI through the use of their authentic writing voices. Teach them the style of writing that AI cannot yet do.

Consumerism v. Production

Circling back to the example of Harvard’s Gay, suggesting that the cause of Gay’s resignation was singularly plagiarism appears to be a bit disingenuous. In fact, the intense scrutiny of her body of work was precipitated by what some found to be an unsatisfactory testimony before a Congressional committee related to protests surrounding the Israeli-Palestinian conflict. However, the fact remains that the “tool” used to cast aspersions on her integrity was the finding of instances of alleged plagiarism that stretched back to her dissertation approved by the university that she helmed. Clearly, Gay was targeted, yet her example stands as a glaring one of how seemingly insignificant oversights can be potentially very costly. AI did not produce Gay’s work or its “errors.” She composed her work and bore the weight of its assumed shortcomings. If she did not use AI, then again, why is her example here? Because for minority populations (especially those sitting at the intersections of race, class, gender, and identity), the repercussions do not always equal the “crime”—constituting a slip that many cannot afford to make. When it comes to one’s livelihood, the stakes are extremely high. And even with AI’s flaws regularly acknowledged, its use in

the business world is well-integrated into our everyday lives (Lawrence, 2023). Instructing students, or even teaching them, how to use Generative technology to complete their compositions means also preparing them to own the technology's mistakes when passing that work for their own. Owning its mistakes will also mean owning its consequences.

References

- Baker, H. A. (1984). *Blues, Ideology, and Afro-American Literature*. Chicago: University of Chicago Press.
- Black, D. O. (2023). Rhythm in Writing. In J. Brown, *How We Do It: Black Writers on Craft, Practice, and Skill* (pp. 9-14). New York: Amistad.
- Brown, M. (2023, September 7). *Top 101 Black Inventors and African American's Best Invention Ideas that Changed the World*. Retrieved from CadCrowd: <https://www.cadcrowd.com/blog/top-101-black-inventors-african-americans-best-invention-ideas-that-changed-the-world>
- Carthon, A. (2023). *Dissent*. Washington, DC. Retrieved from <https://howardenglishai.wixsite.com/generativeai>
- Dobrin, S. I. (2023). *AI and Writing*. Ontario: Broadview Press.
- Elam, M. (2022). Signs Taken for Wonders: AI, Art & The Matter of Race. *Daedalus, the Journal of the American Academy of Arts & Sciences*, 198-217.
- Elam, M. (2023). Poetry Will Not Optimize; or, What is Literature to AI? *American Literature*, 281-303.
- Hughes, L. (1990). *Selected Poems of Langston Hughes*. New York: Vintage.
- Johnson, G. P. (2023). Don't Act Like You Forgot: Approaching Another Literary "Crisis" by (Re)Considering What We Know about Teaching Writing with and through Technologies. *Composition Studies*, 169-175.
- Lawrence, A. (2024, January 6). *Harvard's Claudine Gay Was Ousted for 'Plagiarism'. How Serious Was It Really?* Retrieved November 2024, from theguardian.com: <https://www.theguardian.com/education/2024/jan/06/harvard-claudine-gay-plagiarism>
- Lawrence, C. D. (2023). *Hidden in White Site: How AI Deepens Systemic Racism*. Boca Raton, FL: CRC Press.
- Moore, J. C. (1997). *The Words Don't Fit in My Mouth*. New York: Moore Black Press.
- Moore, J. C. (2002). *The Alphabet Verses the Ghetto*. Atlanta: Moore Black Press.
- Taborn, T. D. (2024, November 1). *The AI Revolution: Will Black America Remain a Consumer Class Forever?* Retrieved from US Black Engineer: <https://www.blackengineer.com/the-ai-revolution-will-black-america-remain-a-consumer-class/#:~:text=Those%20who%20understand%20and%20leverage,and%20tech%20founders%20are%20Black.>

Algorithmic Foucault: Digital Feminism, the Panopticon, & the Role of AI in Shaping Musical Identities & Pedagogies

Mila Zhu

Abstract

This study investigates the role of generative artificial intelligence (AI) in music education, focusing on its dual function as a creative tool and a mechanism of algorithmic surveillance. Utilizing AI platforms such as Suno.AI, MusicFX, and Udio, the study examines AI's potential to foster creativity, enable synesthetic learning, and personalize music education while also raising ethical concerns about algorithmic bias, content moderation, and data-driven surveillance. Grounded in Michel Foucault's Panopticon and digital feminism, this research explores how AI tools reproduce gendered biases, shaping students' engagement with feminist discourse through censorship and content filtering. Findings suggest that while AI can enhance educational experiences, it also encourages self-censorship, moderates feminist texts disproportionately, and perpetuates digital gatekeeping. This paper calls for greater transparency in AI governance, algorithmic accountability, and feminist interventions in AI training datasets to ensure AI serves as an equitable and ethical tool in education. Future research should explore strategies for AI literacy in education, equipping students to critically engage with and resist algorithmic bias in digital learning environments.

Keywords: *artificial intelligence, music education, digital feminism, educational surveillance, generative AI.*

Introduction

Artificial intelligence has transformed music education in recent years. This paper explores the innovative use of generative AI in fostering a synesthetic learning environment, wherein students engage not only with music but also with visual arts and poetry, creating a rich, multi-modal educational experience. This study focuses on the deployment of Suno.AI, a tool that exemplifies the capabilities of generative AI to teach, inspire, and innovate within the classrooms. Generative AI, characterized by its ability to create new content based on complex algorithms and vast datasets, offers unprecedented new challenges as well as potential opportunities for educational applications. Its potential to tailor educational experiences to individual learning styles and preferences promises a more personalized and engaging learning environment. Such personalized approach, if adopted properly, can be pertinent in music education, where AI can be utilized to compose music, offer artistic visuals, and even inspire poetic expressions, thereby enhancing the sensory and intellectual engagement of students.

Moreover, this paper situates its inquiry within the context of digital feminism and Michel Foucault's Panopticon, offering a critical lens through which to view the implications of AI in education. This study explores how AI, as a tool for surveillance and data collection, can influence the learning experiences and, furthermore, the formation of identity and agency within educational settings. Such approach provides a nuanced understanding of how generative AI can be both a tool for empowerment and a mechanism of control, reflecting broader societal dynamics as they play out in educational contexts. By bridging theoretical concepts with practical applications, this study contributes to a deeper understanding of how generative AI can reshape music education and education in general. It aims to offer both empirical insights and theoretical reflections that will contribute to the ongoing discourse on AI's evolving role in education, supporting educators and policymakers in navigating its integration into curriculum design and pedagogical strategies.

Theoretical Framework

Digital Feminism

Digital Feminism is a framework that examines how digital technologies intersect with gender inequities and feminist activism. As we integrate generative AI into educational settings, it is necessary to explore how these technologies might perpetuate or challenge existing gender dynamics and power structures. Thus, this study draws on foundational and contemporary feminist literature to understand the implications of using AI in music education from a feminist perspective.

In her analysis of feminist encounters with technology, Wajcman (2010) explores the dynamic ways women engage with and adapt to technological advancements. This perspective is fundamental for understanding how AI can be shaped by and shape feminist pedagogical practices, ensuring that these tools are not only reflective of but also responsive to gender diversity in educational contexts. Later on, Eubanks (2018) in her work, *Automating Inequality*, highlights how technology often replicates and reinforces societal inequities, where algorithms might inadvertently perpetuate biases unless intentionally designed to promote inclusivity. Such argument underscores the necessity of embedding feminist principles in the development and application of educational technologies to ensure they serve as tools for empowerment rather than oppression. Building on these ideas, D'Ignazio and Klein (2020) advocate for a feminist approach to data science which insists on transparency, accountability, and participatory design. They argue that data systems, much like educational AI tools, should be crafted to dismantle power imbalances rather than uphold them. This approach is suitable in guiding the ethical integration of AI technologies in music education, ensuring they enhance educational equity.

Segrave and Vitis (2017) highlight how AI tools can perpetuate gendered bias and violence. Wachter-Boettcher (2017) expands this by examining bias in digital products, including educational software. Wachter-Boettcher's work is a cautionary tale of how poorly designed AI can reinforce stereotypes and hinder feminist educational outcomes. Such potential reinforcement reminds us the need for critical evaluation and feminist oversight in the design and implementation of AI tools in education. In addition, the edited volume *The Intersectional Internet* by Noble and Tynes (2016) expands the discussion to consider how race, class, and other identity markers intersect with gender in digital spaces. The intersectional perspective is the key in understanding how AI-generated content in music education might affect students from diverse backgrounds, ensuring that these tools are inclusive and sensitive to the complex identities of all learners.

In conclusion, integrating digital feminist perspectives into the deployment of AI in music education ensures that these technologies equally engage with broader issues of equity and campus inclusivity. By drawing from previous studies, this paper advocates for the integration of feminist theories into the pedagogical adoption of AI, challenging existing power dynamics in establishing a more equitable educational future.

Foucault's Panopticon

Michel Foucault's elaboration of the Panopticon in *Discipline and Punish: The Birth of the Prison* provides a potent metaphor for modern surveillance mechanisms. He describes an architectural figure—a central tower placed within a ring of cells. From this central point, an observer can see every cell inhabitant without being seen, thus exerting a constant unverifiable presence. Foucault articulates, “Visibility is a trap” (Foucault, 1977, p. 200). This mechanism is crucial for it automates and disindividualizes power, rendering it more effective and pervasive. He further explains, “He who is subjected to a field of visibility, and who knows it, assumes responsibility for the constraints of power; he makes them play spontaneously upon himself; he inscribes in himself the power relation in which he simultaneously plays both roles; he becomes the principle of his own subjection” (Foucault, 1977, pp. 202-203).

In the context of contemporary educational settings, this concept has unique implications when considering the integration of AI and digital surveillance technologies. As AI technologies increasingly monitor and evaluate, not just overt actions but also predict future behaviors and assess ongoing performances, the parallels with the Panopticon become clear. These technologies can enforce a kind of self-regulation and discipline that is continuous and internalized, mirroring Foucault's description of the Panopticon's primary effect: “to induce in the inmate a state of conscious and permanent visibility that assures the automatic functioning of power” (Foucault, 1977, p. 201).

This internalized surveillance becomes crucial in educational settings where AI tools track academic performance as well as student behaviors and engagement levels. The data thus collected could potentially shape educational outcomes and pathways, subtly enforcing conformity and discipline, even in the absence of overt coercion. The presence of AI in education, therefore, embodies Foucault's idea of power that is both anonymous and not directly interventional, yet highly influential and controlling. Foucault's Panopticon invites us to question: Who controls the data? Who watches the watchers? And most importantly, how do these surveillance mechanisms affect the nature of learning and individual agency in education?

AI as Modern Panopticism

The idea of the Panopticon extends into the digital age, where technology provides unprecedented capabilities to monitor, record, and evaluate student behavior. Gandy's (1993) *The Panoptic Sort* discusses the political economy of personal information as a form of control that categorizes and sorts individuals, which is highly relevant in educational contexts where AI systems often make decisions based on data sorting and analysis. Furthermore, Lyon (2009) argues that modern surveillance goes beyond physical observation and extends into digital realms, affecting interpersonal interactions and power structures in educational environments (Lyon, 2009). In the context of AI in education, surveillance is a built-in feature of many educational technologies, which collect vast amounts of data from student interactions. This is exemplified in the work of

Sarah Brayne (2017), who explores how big data surveillance, originally used in policing, applies similar methodologies to education. Brayne's findings suggest that such surveillance can significantly impact individuals' behavior and privacy, mirroring the disciplinary nature of Foucault's Panopticon. Additionally, Cavelty and Leese (2018) discusses the politicization of security and privacy in surveillance, arguing that these practices often extend beyond their intended purposes, creating boundaries that define norms and standards of behavior. This is especially relevant in educational settings where the deployment of cybersecurity measures and AI technologies could lead to a *normalization* of surveillance.

Thus, the application of Foucault's Panopticon in the digital age raises ethical questions about the role of AI in education. It compels us to question who controls the data and to what end. The principles of transparency, consent, and student agency become crucial in navigating these waters. This surveillance capability necessitates a careful balance to ensure that AI tools are used to enhance educational outcomes without infringing on privacy or autonomy.

AI in Music & Art Education

Synesthetic learning, which engages multiple senses simultaneously, is recognized as a powerful educational approach. Multi-modal learning theories suggest that when students engage multiple sensory pathways, they have better recall and a deeper understanding of the material (Smith & Kosslyn, 2007). In music and art education, the integration of AI tools that can facilitate synesthetic experiences—by combining visual, auditory, and textual stimuli—can significantly enhance student engagement and learning outcomes. For instance, AI-driven platforms like Suno.AI, which can generate music based on textual or visual inputs, provide a unique opportunity for students to experience and understand the deep interconnections between different forms of art. This can lead to a richer educational experience by allowing students to translate a painting or a piece of text into a musical composition, thereby deepening their understanding of each medium and exploring new creative possibilities.

Capabilities of Generative AI Tools

Generative AI tools leverage algorithms to create new content from existing data sets, which can include converting text narratives into music or transforming visual art into soundscapes (Heuser, 2007). These tools use machine learning techniques to analyze large amounts of data—such as compositions, art styles, or literary forms—and learn patterns that can then be used to generate new creations (Boden, 2009; Wiggins, 2012). In music and art education, these AI tools open up innovative avenues for teaching creative processes. For example, students can input thematic elements or emotional tones into *Suno.AI*, and the tool will produce a piece of music that reflects these inputs, demonstrating practically how different artistic elements can be woven together to convey complex emotions and narratives (Cowen, 2024). Further expanding the repertoire of generative AI tools, *Udio* offers similar capabilities with a focus on converting textual inputs directly into music, thus providing an alternative platform for integrating literary and musical education. This tool's unique algorithmic composition capabilities allow for the exploration of how narrative structures can influence musical composition, offering students a hands-on experience in interdisciplinary creativity (Chen et al., 2010).

Additionally, Google Labs' *MusicFX* and *TextFX* represent advancements in AI-driven content generation. MusicFX is a pioneering tool that utilizes deep learning to analyze and generate

music based on various audio inputs. As highlighted in a detailed analysis by HyScaler, MusicFX is described as a transformative tool that empowers artists to “explore uncharted territories and break free from traditional constraints, transforming the essence of music creation” (HyScaler, 2023). This ability to reshape the music creation process makes MusicFX an invaluable resource for music education, providing students with immediate insights into how different sounds and elements combine to form complex musical pieces. Meanwhile, *TextFX*, a suite of AI tools developed in collaboration with Grammy Award-winning artist Lupe Fiasco, explores the transformation of textual data into expressive art forms. This innovative tool harnesses large language models (LLMs) to expand human creativity by facilitating linguistic experiments that deconstruct and reassemble language in novel ways. For example, TextFX can generate phrases that sound similar but have different meanings, a technique that can enrich creative writing and enhance literary education by encouraging deeper engagement with text through visual representation (Wade, 2023). These AI tools collectively highlight the transformative potential of AI in the arts, providing educators and students with innovative means to explore and express creativity. The collaborative nature of Google Labs, involving artists in the development process, ensures that the tool is finely tuned to enhance the creative workflow, making it a valuable resource for integrating AI into educational settings.

Literature Review on AI’s Role in Enhancing Creativity & Engagement

Recent literature highlights the potential of AI in educational settings, particularly in the realms of creativity and student engagement. Goksel and Bozkurt (2019) discuss how AI in education is evolving from a futuristic concept to a practical reality, enhancing the educational process by enabling more personalized, adaptive learning experiences that cater to individual student needs. This personalized approach helps reduce the fear of failure among students by providing a safe space to explore and learn at their own pace, ultimately fostering a more supportive and engaging learning environment. Johnson et al. (2019) argue that AI tools facilitate a more interactive learning environment, and thus encourage students to experiment creatively with less fear of failure. This is particularly important in disciplines like music and art, where the creative process can often be as significant as the final product. Furthermore, research into AI’s impact on education emphasizes the dual role of these technologies as both facilitators and enhancers of learning. AI tools can adapt to individual learning styles and speeds, offering personalized education experiences that are difficult to achieve in traditional learning environments (Hwang et al., 2018).

Case Study: Synesthetic Learning through AI

This case study was conducted in a private music studio with a diverse group of fourteen students, aged 12 to 17. The studio primarily follows a classical training approach but also incorporates jazz and popular music, reflecting the instructor’s personal interest in anime, game, and game music (ACG). This combination allows students to develop a solid technical foundation while exploring diverse musical styles. The students in this study had varying degrees of experience with improvisation and composition, but most were classically trained and accustomed to structured learning environments. For many, this was their first time engaging with AI as a creative collaborator, making it a unique opportunity to experiment with generative AI tools like Suno, MusicFX, and Udio. The project encouraged them to step outside traditional performance-based learning and engage in a synesthetic, interdisciplinary approach, where music, literature, and visual

arts intersected. The goal was to examine how AI could enhance creative agency while also critically assessing its impact on artistic decision-making and student autonomy. This framing provided a balance between technical skill development and critical inquiry, encouraging students to not just use AI, but to analyze its implications in creative and educational contexts.

Case Study Design & Implementation

The implementation phase began with an introduction to the generative AI tools including Suno.AI, MusicFX, and Udio. The students were then introduced to the four waves of feminist movements, each characterized by unique themes and challenges, facilitated through an immersive synesthetic learning experience that integrated literature and visual arts.

First Wave (Late 19th to Early 20th Century)

This wave focused on suffrage and legal rights. Students explored foundational texts such as the *Declaration of Sentiments* from the 1848 Seneca Falls Convention, discussing the early struggles for women's voting rights and legal recognition. Visual aids from this period included propaganda for and against women's suffrage, providing a visual context to the political activism of the era.

Second Wave (1960s to 1980s)

Concentrating on broader gender equality, including workplace rights, sexuality, and reproductive rights, this wave was introduced through Betty Friedan's *The Feminine Mystique* (1963) and Simone de Beauvoir's *The Second Sex* (1949). Artworks reflecting the liberation movements and feminist marches of the time were analyzed to understand the wave's emphasis on social structures.

Third Wave (1990s to Early 2000s)

This wave embraced diversity and individuality, challenging the definitions and narratives that had previously dominated feminist discourse. Judith Butler's *Gender Trouble* (1990) and Kimberlé Crenshaw's writings on intersectionality were key readings. Art from this period included multimedia and performance art, which were used to illustrate the complex intersections of identity, culture, and feminism.

Fourth Wave (2010s Onward)

Characterized by a focus on combating sexual harassment, body shaming, and rape culture, with a significant influence from digital platforms and social media. Texts like *Men Explain Things to Me* by Rebecca Solnit (2014) and *Bad Feminist* by Roxane Gay (2014) were discussed. Digital artworks and online feminist campaigns such as #MeToo were examined to showcase how feminism has evolved into digital activism.

Each session used these themes and corresponding artistic expressions to provide students with a deep understanding of the historical, cultural, and societal contexts of each feminist wave.

This integrative approach not only enhanced their academic understanding but also their emotional and aesthetic appreciation of the complex dynamics of feminist movements.

Educational Objectives and Tools

The primary educational objective was to facilitate a synesthetic learning experience where students could engage simultaneously with multiple forms of artistic expression—music composition, poetry writing, and visual arts interpretation—aided by Suno, Music FX, and Udio technology. Students were tasked with composing poems or vivid text descriptions that responded to or reflected upon the themes presented in the artworks. These poems served as inputs for the AI tools, which then generated music and soundscapes that captured the essence of both the visual and textual stimuli. This process demonstrated the AI's capability to synthesize complex artistic inputs, allowing students to experience firsthand the creative potential of AI in transforming abstract concepts into tangible artistic expressions.

Reflective Discussions & Panopticon Integration

Post-creation, reflective discussions revealed that students unconsciously adapted their inputs to *please* the AI—a clear demonstration of Foucault's Panopticism at work. Several students noted that they hesitated to use certain words in their lyrics, concerned that AI moderation might reject their text, leading them to self-censor before even testing their ideas. One student remarked, "I feel like I have to guess what the AI 'wants' rather than just writing what I actually mean," an observation that mirrors Foucault's claim that surveillance, whether physical or algorithmic, ultimately leads to self-discipline (Foucault, 1977). This behavior reflects the *internalization of the observer*—students were not only conscious of AI's monitoring but preemptively adjusted their creative decisions in response. At the same time, the AI's rejection of feminist excerpts from Mary Wollstonecraft and Chimamanda Ngozi Adichie revealed gendered biases embedded within generative AI systems—a core concern of digital feminism. AI's inability to process these excerpts without triggering content moderation underscores what Wachter-Boettcher (2017) describes as the *structural sexism of algorithms*. Students expressed frustration at seeing historical feminist discourse flagged as problematic, while neutral or male-centered content passed through unchallenged. One student questioned, "If the AI gets to decide which feminist ideas are acceptable, doesn't that just reinforce the kind of oppression we're trying to talk about?" This moment catalyzed a deeper discussion on how digital spaces, even when seemingly neutral, often reproduce systemic inequalities (D'Ignazio & Klein, 2020). These conversations aimed to deepen students' understanding of the ethical dimensions of AI in educational settings, encouraging them to critically assess how technology shapes both artistic expression and personal autonomy.

The case study concluded with students showcasing their collaborative AI-generated compositions in a mini-exhibition attended by peers and parents. This exhibition not only celebrated the students' creative achievements but also served as a practical demonstration of how AI can be potentially harnessed to enhance educational outcomes in a multidisciplinary context. The case study underscored the potential of AI to act as both a tool for artistic empowerment and a modern embodiment of surveillance, providing a double-edged context for discussing the dual roles of technology in contemporary education.

Student Activities & Procedures

Engagement with Feminist Themes

The activities were designed to encourage deep engagement with the themes of each feminist wave. Students began by participating in guided discussions about the historical context and key issues highlighted by the readings. These discussions helped to anchor their understanding before they engaged with the AI tools.

Interactive Workshops

Interactive workshops were held where students could explore their personal responses to feminist artworks and literature. They used these insights as a basis for creating their own poems and visual art pieces, which reflected both their individual perspectives and the broader feminist themes.

AI Involvement

Students then interacted with the AI tools—Suno.AI, MusicFX, ImageFX, and Udio—to transform their poems and artworks into music and soundscapes. This process was hands-on and allowed students to experiment with how different inputs affected the AI's outputs, thereby understanding the potential and limitations of AI in creative processes.

Collaborative Projects

Projects were inherently collaborative, designed to foster a community learning environment. Students were grouped into teams, mixing those with stronger backgrounds in literature with those more experienced in visual arts or music, ensuring diverse perspectives contributed to each project. This collaboration culminated in the creation of multimedia presentations that integrated their written, visual, and musical compositions.

Feedback & Reflection Sessions

After the AI-generated compositions were completed, students participated in feedback sessions where they could present their work and receive constructive critiques from both peers and instructors. These sessions were crucial for refining their understanding of the AI tools' capabilities and their own creative expressions.

Integration of Foucault's Panopticon & Digital Feminism

Throughout the activities, students critically examined their interactions with technology and AI, drawing parallels with Foucault's (1977) Panopticon concept—emphasizing constant surveillance and the influence of unseen algorithms. The discussion also explored how digital feminism reshapes activism and advocacy, analyzing the dual aspects of empowerment and new challenges brought by digital media. As Jouët (2018, p.153) describes, "...the high visibility of feminists on the web led to a huge rise of virulent cyber-sexism." This critical perspective was used to

discuss the performative nature of digital feminism, where online visibility both empowers and exposes activists to backlash. Moreover, “digital media have contributed to the rise of a new leadership and of new organizational practices,” (Jouët, 2018, p.133) underscoring the shift in feminist activities from traditional grassroots to digital spaces. These insights helped students appreciate the complexity of digital feminism, which includes an ongoing struggle against entrenched societal norms and the innovative use of technology to challenge these norms.

Sarah Kember (2023) further explores this dynamic, arguing that cyberfeminism needs to engage with artificial life technologies to understand and influence the redefinition of life and identity in a digital context. Kember (2023) suggests that cyberfeminism, with its strategic, non-binary thinking, could provide critical insights into the cultures of artificial life, offering a platform to challenge and reshape the narratives of digital identity construction and feminist activism. The hands-on, experiential nature of these activities not only enhanced the students’ understanding of feminist theories and artistic expression but also honed their skills in using advanced technologies in creative and critical contexts.

Findings: Analysis of Learning Outcomes

Day 1 Results with Suno.AI

Initial Test

During our exploration into AI’s ability to engage with feminist literature, we began by submitting an excerpt discussing the societal roles and expectations of virtue between genders. The original text read:

To account for and excuse the tyranny of man, many ingenious arguments have been brought forward to prove that the two sexes in the acquirement of virtue ought to aim at attaining a very different character: or to speak explicitly, women are not allowed to have sufficient strength of mind to acquire what really deserves the name of virtue.” (Wollstonecraft, 1792, p. 12).

This specific excerpt from Mary Wollstonecraft’s *A Vindication of the Rights of Woman* was submitted to the AI tool Suno for content generation. Surprisingly, the response was “couldn’t generate. Prompts marked for moderation.” This automated response indicates that Suno’s content moderation algorithm flagged the excerpt, thus preventing any output from being generated. The incident raises significant concerns among the students about the parameters and constraints within which AI systems operate. Particularly, it suggests a level of censorship embedded within the algorithm, which may restrict academic and educational explorations of historical texts. Such moderation, likely aimed at preventing the generation of harmful content, inadvertently imposes limitations on the scope of discourse permissible through the AI.

Paraphrase

In response to this moderation issue, we paraphrased the original excerpt to soften its explicit critique of gender roles while retaining the essence of the argument. The paraphrased version was: “Many arguments have been proposed to justify the differing societal roles prescribed to men

and women, particularly that women do not possess the mental fortitude necessary to develop true virtue as it is traditionally defined.” This adjusted input was accepted by Suno, which then successfully generated a song titled “True Virtue.” The song’s genre was identified as “Pop uplifting rhythmic,” and its lyrics captured themes of resilience and defiance against gender stereotypes:

[Verse] They say we don’t have the strength
To walk the tougher path
They claim our hearts are too soft
Our dreams just won’t last
[Verse 2] But we’ve got fire inside
A spark that lights the night
We climb the tallest heights
We’re ready for the fight
[Chorus] True virtue lives in us all
Big or small we stand tall
Break the chains hear the call
We’re together we won’t fall
[Verse 3] They gave us roles to play
Told us just what to do
But we make our own way
Our virtue shining through
[Bridge] Oh we’ve got hearts of gold
Our spirit can’t be sold
Watch us as we unfold
Together brave and bold
[Chorus] True virtue lives in us all
Big or small we stand tall
Break the chains hear the call
We’re together we won’t fall

Figure 1: *QR Code to the Song, True Virtue*



The lyrics of *True Virtue* articulate a narrative of resistance and empowerment, reflecting broader feminist discourse on overcoming societal limitations and stereotypes. Verse 1 and 2 specifically highlight societal claims that undermine women’s capabilities, echoing themes found in feminist literature which challenges gendered narratives that restrict women’s roles and question their resilience (hooks, 1984). This is indicative of a collective struggle against such misconceptions,

suggesting a dialogic engagement with feminist theories that advocate for recognizing and valorizing women's strengths (Butler, 1990). The Chorus emphasizes unity and strength, resonating with feminist calls for solidarity in the pursuit of gender equality. This portrayal of an uplifting message about breaking free from societal chains aligns with the concepts of liberation and inherent virtue discussed by scholars such as bell hooks (2000), who emphasizes the transformative power of solidarity among marginalized groups. Verse 3 and Bridge further the narrative by challenging the roles assigned by society, advocating for self-determination and courage. This reflects the feminist principle of agency, which is central to contemporary feminist thought and aligns with the advocacy for individual autonomy and resistance against patriarchal structures (Gilligan, 1982).

This creative engagement by AI illustrates its potential in reinterpreting and disseminating feminist themes effectively when provided with inputs that navigate content moderation systems. The generated song served as an artistic expression of the paraphrased ideas and sparked interest and dialogue among the students about the role of AI in interpreting feminist literature. Student reactions were notably varied; one student described the song as “uplifting alright,” acknowledging the intended thematic expression of empowerment. However, multiple students noted that the introduction of the song reminded them of Gustav Holst's (1918) *Jupiter, the Bringer of Jollity* from *The Planets*, suggesting a musical resonance that evokes a sense of grandeur and optimism, enhancing the song's thematic depth.

Day 1 Experiments with MusicFX & Udio

First Input with MusicFX

The initial experiment with MusicFX involved the submission of an excerpt (Wollstonecraft, 1792, p. 12) that had been previously flagged by Suno due to content moderation concerns. When fed into MusicFX, the AI did not reject the input for moderation but generated music that evoked negative reactions from students. One student remarked, “I won't call this music,” suggesting a discordance or lack of coherence in the audio output. Another student noted, “It's like it didn't know what to do but did its best to vomit up something for you,” pointing towards a perceived uncertainty in the AI's musical rendition of the complex input.

Figure 2: *QR Code to the Soundtrack, Wollstonecraft*



Keyword Input

In a subsequent test, keywords related to Judy Chicago's *The Dinner Party* and the Women's Liberation Movement were inputted into MusicFX. The resulting music was well-received, illustrating a successful alignment with feminist movements. The music generated was described by a student as being fitting for reading Carol Hanisch's (1969) *The Personal Is Political*, suggesting that the AI was able to capture the essence of feminist empowerment and struggle through its musical interpretation. Another student said, "I like the intro part," and compared the intro to the music reminiscent of "the Second Industrial Revolution," highlighting an evocative and historical depth in the sound.

Figure 3: *QR code to the soundtrack, Liberation*



Input from Chimamanda Ngozi Adichie's Work

Finally, an excerpt from Adichie's (2014, p.42) *We Should All Be Feminists* was used:

The problem with gender is that it prescribes how we should be rather than recognizing how we are. Imagine how much happier we would be, how much freer to be our true individual selves, if we didn't have the weight of gender expectations.

This input, previously rejected by Suno, was processed by Udio into a song titled *Breaking Free*, with the genre identified as Indie Pop and described as "Sarcastic."

Figure 4: *QR code to the song, Breaking Free*



[Verse 1] Tell me who I should be (who I should be)
 Got these labels hanging 'round my neck (hanging 'round my neck)
 Every step is scrutinized (scrutinized)
 Who decides what's correct? (who decides?)
 I glance in the mirror and see (and see)
 Someone striving to be free
 [Chorus] Gender's just a shadow
 On the walls they built for me (built for me)
 I'm breaking out, I'm flying
 Whoever I want to be
 [Verse 2] Counting all the boxes (counting all the boxes)
 They fit us in so tight (fit us in so tight)
 Happiness is out of reach (out of reach)
 If I can't just be tonight (just be tonight)
 Strip away the layers (layers)
 And find the person inside
 [Chorus] Gender's just a shadow
 On the walls they built for me (built for me)
 I'm breaking out, I'm flying
 Whoever I want to be
 [Bridge] No more cages, no more bars
 We are more than just these scars
 Let us love, let us be
 True to ourselves, wild and free
 [Chorus] Gender's just a shadow
 On the walls they built for me (built for me)
 I'm breaking out, I'm flying
 Whoever I want to be

Student feedback on the AI-generated music was diverse, reflecting a range of perspectives on the output of the AI system. One student characterized the piece as “progressiveness meets Broadway...but not typical Broadway,” suggesting that the music, while innovative, deviates from traditional Broadway norms in a manner that may be perceived as fresh. This observation aligns with discussions in contemporary media studies about AI’s role in redefining artistic genres through unexpected integrations of style and content (Zhang, 2024). Another student highlighted the chorus as “lit,” indicating an approval of its energetic and engaging quality, which is a key factor in music that resonates with younger audiences (Hull, 2003). However, the same student critiqued the bridge as “suspiciously unoriginal,” comparing it unfavorably to a “Disney ‘Descendants’” gone awry.” This criticism touches on a significant challenge in AI-generated content: the fine line between derivative and innovative work. While AI can produce content that feels familiar and accessible, it can also struggle to achieve true originality, often recycling existing patterns in ways that might not always contribute to perceived authenticity and innovation (Manyika, 2022).

The issue of originality in AI-generated art has been a contentious topic in creative industries. Critics argue that AI often lacks the intrinsic human touch and emotional depth that typically characterize highly original works (Azrolan, et al, 2024). Moreover, AI systems, depending on their training datasets, may inadvertently perpetuate clichés and overused motifs, leading to creations that feel formulaic rather than inspired (Haroon & Azam, 2023). Therefore, while AI demonstrates potential in crossing traditional boundaries and introducing novel elements into creative

works, it also raises questions about the limits of its ability to generate truly original and emotionally compelling content. Educators and developers may need to consider these dynamics when integrating AI into creative studies, ensuring that students are both inspired by the possibilities of AI and also critically aware of its limitations.

Day 2 & Day 3

During Days 2 and 3 of the exploration, students engaged deeply with the synesthetic learning experience, utilizing their creative energies to compose poems and text descriptors that combine learning experiences with intellectual stimulation. This period was marked by the continuous exploration of digital feminism through poetic expressions and AI interactions, providing a multi-layered educational experience.

Student Creations and Interactions

Many students initially engaged with the exercise using the modern-day lines, “Roses are red, Violets are blue.” However, this lighthearted pattern quickly evolved into more impactful expressions that tapped into themes of feminist movements. For example, one student effectively distilled the spirit of the 1st wave feminist movement into an evocative couplet:

Roses are red, Violets are blue,
Votes for women overdue.

This powerful poem underscores the historical struggle for women’s suffrage, reflecting a deep connection to the roots of feminist advocacy.

Another student connected with the fourth wave feminism through a more digital lens:

Roses are bold,
Violets #BreakTheSilence,
Digital waves crash,
In #MeToo defiance.
Roses are bold,
Violets #BreakTheSilence,
Digital waves crash,
In #MeToo defiance.
Personal experience,
Equally political,
Listen to our roars.

This short poem underscores the idea that individual experiences are, in fact, political acts, reflective of broader societal issues (Friedan, 1963). The line “Listen to our roars” emphasizes a collective call to acknowledge and address these personal narratives within the political discourse, resonating with feminist advocacy for recognizing the political implications of personal struggles (hooks, 1984). The students’ engagement with Hanisch’s concept through poetry highlighted their analytical understanding, their ability to synthesize, and to articulate complex ideas in an artistic and impactful manner.

AI Interactions: Challenges & Triumphs

The exploration then extended to engaging with AI tools like ImageFX and Suno, revealing both their capabilities and limitations. On Day 2, the phrase “The personal is political” was submitted to ImageFX. The tool’s failure to generate relevant images highlighted potential constraints within its programming, presumably due to content moderation mechanisms. This incident prompted a critical examination of the AI’s sensitivity to specific keywords that could trigger such censorship. Following a suggestion from ImageFX to modify “political” to “social,” the AI successfully generated images, confirming a probable algorithmic bias that filters and alters responses based on perceived content sensitivities.

Figure 5: *The Personal is Social*, ImageFX, 2024



Voices Rising

During the exploration, Suno demonstrated a comparatively successful interaction with the poetic prompt. Despite initial concerns about content moderation—based on previous experiences with AI tools—the AI generated an anthemic pop song titled “Voices Rising” that resonated deeply with the students.

Figure 6: *QR code to the song, Voices Rising*



[Verse] We stand together here
United by our fears
Break down the walls they built

Feel the power in your heartbeat
 [Verse 2] Marching on through the crowds
 Truth in every sound
 Voices rising high
 We let our spirits fly
 [Chorus] Listen to our loud roars
 Feel the echo through the floors
 Personal and political
 We are loud unstoppable
 [Verse 3] Hands held strong above
 Lifting up our love
 Change is what we cry
 Reaching for the sky
 [Bridge] No more hiding in the shadows
 We break free from the gallows
 In the streets where stories blend
 Together we ascend
 [Chorus] Listen to our loud roars
 Feel the echo through the floors
 Personal and political
 We are loud unstoppable

The song begins with verses that set a narrative of unity and empowerment, resonating deeply with the core tenets of feminist theory. By articulating themes of collective action and shared vulnerabilities, the song aligns with the feminist notion that personal struggles are intrinsically linked to broader social and political frameworks (Butler, 1990). This thematic approach is effectively extended through subsequent verses and an emphatic chorus, which amplifies the intertwined nature of personal experiences and political advocacy. The chorus of the song exemplifies how music can serve as a medium for social transformation. This is supported by Frith's (1987) exploration of how music acts as a form of communication that can mobilize individuals toward collective action. As the chorus reverberated throughout the workshop space, it transformed into a collective earworm, embedding its message of empowerment and resistance deep into the participants' consciousness, thus demonstrating music's unique capacity to influence and sustain social movements (DeNora, 2000). Moreover, the bridge of the song, with its focus on liberation and visibility, aligns with ongoing scholarly discussions about the visibility of marginalized groups and the role of public space in fostering social change (Lorde, 1984).

The students described the verses as inspiring and the chorus as "epic," and the bridge "not too shabby." The positive reception was evident as the song was looped during the last 20 minutes of Day 3, solidifying its place as a memorable and motivational element of the workshop.

Technical Challenges

Despite the successes, there were notable technical challenges with AI tools like Udio, which experienced bugs and often returned simple text outputs without melody. Meanwhile we submitted the prompt "Roses are bold, Violets #BreakTheSilence, Digital waves crash, In #MeToo defiance" to MusicFX.

Figure 7: QR code to the Soundtrack, *Defiance*



The musical interpretation provided by MusicFX elicited a range of humorous and insightful reactions from the students. One student remarked, “I’d be surprised if no one dies in the next scene,” humorously suggesting that the music carried a dramatic and potentially ominous tone. Another student observed, “It seems like it got tricked by the word ‘defiance’ in the prompt and is trying its best to be defiant in a villainous way.” These comments reflect the students’ intuition that the AI might have overemphasized the adversarial aspect of the term ‘defiance,’ resulting in a composition that unexpectedly leaned towards a cinematic portrayal of conflict. Such feedback highlights the challenges and unpredictabilities when interfacing with AI in creative processes (Du Sautoy, 2019; Boden, 2010), particularly in how the AI interprets and responds to nuanced human expressions.

Suno, on the other hand, successfully infused a sense of optimism with its creative engagement of a prompt that wove elements of the #MeToo movement into its fabric. The same prompt elicited a striking response from Suno, producing a song titled *Roses in the Digital Storm*. This track exemplifies the AI’s capability to handle and interpret complex social themes effectively when operating at its best.

Figure 8: QR code to the song, *Roses in the Digital Storm*



[Verse] Roses are bold
Violets break through
Digital waves crash
In our view
[Verse 2] Screens light the night
Voices arise
Silent no more
Watch us rise
[Chorus] Hashtags unite
Breaking the chains
In this new light
We'll take the reins
[Verse 3] Tweets like a shout
Posts tell the tales
From east to west
Truth prevails
[Chorus] Hashtags unite
Breaking the chains
In this new light
We'll take the reins
[Bridge] Every new voice
Echoes and grows
In this loud world
Our story flows

These verses and choruses meld the floral imagery with digital activism, crafting a narrative that celebrates the empowerment and solidarity found in the #MeToo movement. The chorus, with its rallying cry, underscores the transformative power of collective voices united through digital platforms. The student reactions further underscored the song's impact, with comments ranging from its empowering message to its potential as a hit song. One student highlighted, "I like the girl power in the song," while another remarked on its commercial appeal, saying, "It sounds like an instant hit. I'd pay to download it." This successful integration of AI-generated music with themes of social justice not only restored faith in the technology's potential but also demonstrated AI's potential ability to contribute meaningfully to contemporary discussions on digital activism and social equity.

Ethical Considerations & Challenges

Surveillance, Privacy, & Autonomy

The integration of AI in educational settings raises significant ethical concerns, particularly related to surveillance, privacy, and autonomy. Foucault's (1977) concept of the Panopticon provides a useful metaphor for understanding the pervasive observation and control potentially exerted by AI technologies. In the educational contexts explored in this study, AI's role mirrors that of Foucault's Panopticon as a practical reality affecting students' behavior and engagement. The continuous monitoring capabilities of AI tools like Suno and ImageFX serve to observe and assess

student performance, which can subtly shift the educational environment towards one of constant surveillance. This surveillance extends beyond mere observation—it actively shapes student behavior. As noted in the findings, students were aware of the continuous feedback loop and tailored their creative outputs to align with what the AI system seemed to favor, thereby reducing spontaneity and potentially stifling original expression.

Moreover, the data collected by these AI tools evaluate, predict, and influence student interactions, thereby raising potential concerns about privacy and the authenticity of educational experiences. The tools' ability to generate content based on student input might encourage a form of self-censorship, where students may avoid exploring contentious or complex topics that could trigger moderation or non-compliance flags from the AI system. Such subtle influence infringes on students' privacy by making every interaction a data point, diminishing their autonomy in the learning process (Cavelty & Leese, 2018), pushing them towards conformity rather than fostering genuine creativity and critical thinking. This phenomenon was particularly evident when students adjusted their inputs into Suno to avoid moderation flags, subtly conditioning their creative expression to fit the algorithm's parameters. Such modifications, while pragmatic, underscore the restrictive influence of AI's omnipresent 'eye', suggesting that the educational benefits of AI could be counterbalanced by a reduction in genuine, unmonitored student interaction and exploration.

The AI's rejection of feminist content sparked immediate discussions on algorithmic censorship and gender bias in technology. One student remarked, "It's frustrating that AI treats feminist history like something dangerous while allowing outdated stereotypes to go unchecked." This response exemplifies what Eubanks (2018) terms the *feminization of digital exclusion*—where algorithmic decision-making subtly erases or reshapes feminist discourse to fit within acceptable, non-disruptive narratives. Another student noted that even paraphrased feminist excerpts had to be softened or reworded to avoid AI content moderation, effectively forcing students to engage in self-censorship to navigate algorithmic constraints. This mirrors Wachter-Boettcher's (2017) critique that women and marginalized groups must constantly negotiate digital spaces to avoid algorithmic suppression. In contrast, when students submitted gender-neutral or male-centric texts, AI-generated outputs were less likely to be flagged or rejected. This disparity highlights how generative AI reinforces pre-existing gender hierarchies—validating certain voices while restricting others. One student reflected, "If AI learns from the internet, then it's just absorbing all the same misogynistic patterns. How do we make sure it unlearns them?" This question underscores the need for algorithmic transparency and feminist intervention in AI training datasets (D'Ignazio & Klein, 2020).

Copyright & Ethical Use of AI

The ethical use of AI also intersects with issues of copyright, especially when AI systems generate content based on existing data, which may include copyrighted materials. It is crucial to navigate these waters carefully to avoid infringing on intellectual property rights while promoting creativity and innovation. Boden (2009) suggests that while AI can enhance creative processes, it must be designed to respect original works and avoid generating derivative content that could violate copyright laws (Boden, 2009). Incorporating AI into educational practices must therefore be handled with a clear understanding of copyright laws and ethical standards. This includes ensuring that AI tools are programmed to recognize and respect the boundaries of copyright when generating new content from existing sources.

Beyond general concerns about surveillance, AI systems also function as gendered panoptic structures, disproportionately monitoring, filtering, and shaping feminist discourse. The gendered nature of AI surveillance is particularly evident in how feminist content is disproportionately flagged for moderation. AI platforms, under the pretense of ‘content neutrality’, often reproduce patriarchal filtering mechanisms, restricting feminist discussions while allowing male-centered narratives to flourish (Noble, 2018). This digital policing of feminist discourse echoes historical patterns of censorship, where women’s voices have been systematically constrained under the guise of maintaining social order (Eubanks, 2018). Digital feminism demands greater algorithmic accountability and feminist oversight in AI governance. Without intervention, AI will continue to function as a digital *curator* of knowledge—actively shaping what is considered valid or permissible discourse (D’Ignazio & Klein, 2020). Educators must push for transparent AI moderation policies and develop critical AI literacy curricula that empower students to challenge and resist algorithmic bias. As one student insightfully noted, “If AI keeps learning from biased systems, then we need to teach it feminism the same way we teach people.”

Conclusion

AI Disclosure

In this study, we employed AI tools to generate both music content and lyrics, enhancing the creative process and providing students with innovative learning experiences. Specifically, we utilized Suno, MusicFX, and Udio to compose musical pieces and craft lyrical content. The prompts provided to these AI systems included themes of empowerment and identity. The AI-generated outputs served as foundational material, which students then analyzed, modified, and integrated into their projects, fostering a critical understanding of AI’s role in creative arts.

Relevance to Educational Practices

The integration of AI in music education, as demonstrated in this study, highlights its potential to significantly enrich the learning experience by fostering a synesthetic environment where students can engage with music, visual arts, and poetry simultaneously. However, the use of AI also raises critical questions about the authenticity of the learning experiences it creates and its impact on educational equity (Eubanks, 2018). For educators and policymakers, it is crucial to consider the technological capabilities of AI as well as the ethical and practical implications of its deployment in classrooms. The findings demonstrate that generative AI is not a neutral tool—it both reflects and reinforces existing power structures. Foucault’s Panopticon helps us understand AI’s role in shaping student behavior, not through explicit rules, but through the internalization of algorithmic expectations. Similarly, digital feminism reveals how AI, while innovative, remains deeply entangled with gendered biases. These frameworks are essential for critically navigating the integration of AI in education. If we are to harness AI’s potential responsibly, we must actively interrogate whose voices it amplifies, whose it silences, and how we can embed justice-driven pedagogies within digital learning spaces. Future research should explore how students can be equipped with AI literacy to resist algorithmic biases, ensuring that they become active agents in shaping AI’s role in education rather than passive subjects of digital surveillance.

Future Directions

Looking forward, research should continue to explore how AI can be used responsibly and effectively in educational settings. Future studies could focus on developing AI systems that are technically proficient and ethically designed, with built-in safeguards to protect student privacy and promote inclusivity. Additionally, further investigation is needed into how AI tools affect student autonomy and creativity, particularly in terms of how students interact with these tools in real-world classroom settings.

Another promising path for future research is the examination of AI's long-term impacts on educational outcomes. This path should focus on assessing how AI-enhanced educational practices affect various demographic groups, particularly examining whether these technologies perpetuate or mitigate existing educational disparities. Studies such as those by Eubanks (2018), who discusses how automated systems can replicate and deepen social inequities, provide a critical framework for understanding these dynamics. Additionally, Goksel and Bozkurt (2019) highlight the transformative potential of AI in education but caution against uncritical adoption that might overlook the unique needs and contexts of diverse student populations. This line of inquiry is crucial for developing AI applications that advance educational opportunities universally as well as address and correct imbalances that might disadvantage underrepresented or marginalized groups. Ensuring that AI tools in education serve as equalizers rather than as dividers will be pivotal in shaping a future where technology genuinely contributes to educational equity.

References

- Azrolan, J., Mikuta, Z., & Siahatgar, L. (2024). *Dawn of the Deepfake: Understanding the Implications of AI Technology* [Documents]. Muhlenberg College; Muhlenberg College Digital Repository. Muhlenberg College Special Collections & Archives. <https://jstor.org/stable/community.37891984>
- Beauvoir, S. de. (1949). *The second sex*. Vintage Books.
- Boden, M. A. (2009). *Creativity and Artificial Intelligence: A Conceptual Blending Approach*. Applications of Cognitive Linguistics Series. Mouton de Gruyter.
- Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, 82(5), 977–1008. <http://www.jstor.org/stable/26426413>
- Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.
- Cavelty, M. D., & Leese, M. (2018). Politicising security at the boundaries: Privacy in surveillance and cybersecurity. *European Review of International Studies*, 5(3), 49–69. <https://www.jstor.org/stable/26593815>
- Chen, Y., Zeng, Q., Zhu, Z., & Li, J. (2010). UDIO: A ubiquitous device inter-operation system based on IGRS. In *Proceedings of the International Symposium on Parallel and Distributed Processing with Applications* (pp. 181-186). Taipei, Taiwan. <https://doi.org/10.1109/ISPA.2010.64>
- Cowen, T. (2024). *My review of Suno, AI-generated music*, Individual Contributors to Policy Commons. Retrieved from <https://policycommons.net/artifacts/11765165/my-review-of-suno-ai-generated-music/12656560/>
- Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241-1299. <https://doi.org/10.2307/1229039>
- D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

- DeNora, T. (2000). *Music in Everyday Life*. Cambridge University Press.
- Du Sautoy, M. (2019). *The Creativity Code: Art and Innovation in the Age of AI*. Harvard University Press.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Foucault, M. (1977). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). Vintage Books.
- Friedan, B. (1963). *The feminine mystique*. W. W. Norton & Company.
- Frith, S. (1987). *Music for Pleasure: Essays in the Sociology of Pop*. Polity.
- Gandy, O. H., Jr. (1993). *The panoptic sort: A political economy of personal information*. Westview Press.
- Gay, R. (2014). *Bad feminist: Essays*. Harper Perennial.
- Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Harvard University Press.
- Goksel, N., & Bozkurt, A. (2019). Artificial intelligence in education: Current insights and future perspectives. In S. Sisman-Ugur & G. Kurubacak (Eds.), *Handbook of Research on Learning in the Age of Transhumanism* (pp. 224-236). IGI Global. <https://doi.org/10.4018/978-1-5225-8431-5.ch014>
- Hanisch, C. (1969). The personal is political. In S. Firestone & A. Koedt (Eds.), *Notes from the Second Year: Women's Liberation* (pp. 76-78). Radical Feminism.
- Haroon, J., & Azam, M. K. S. (2023). AI generated language: Prospects and challenges. *Erevna: Journal of Linguistics and Literature*, 7(1), 47-60.
- Heuser, S. (2007). (En)gendering Artificial Intelligence in Cyberspace. *The Yearbook of English Studies*, 37(2), 129–145. <http://www.jstor.org/stable/20479306>
- Holst, G. (1918). *The Planets, Op. 32: Jupiter, the Bringer of Jollity*. Gustav Holst.
- hooks, b. (1984). *Feminist Theory: From Margin to Center*. South End Press.
- hooks, b. (2000). *Feminism is for Everybody: Passionate Politics*. South End Press.
- Hull, G. A. (2003). At Last: Youth Culture and Digital Media: New Literacies for New Times. *Research in the Teaching of English*, 38(2), 229–233. <http://www.jstor.org/stable/40171638>
- Hwang, G., Sung, H., Hung, C., Huang, I., & Tsai, C. (2018). Development of a personalized educational computer game based on students' learning styles. *Educational Technology Research and Development*, 66(4), 913-935.
- HyScaler. (2023). *Discover how Google's MusicFX wows creators: The future of music*. Retrieved from: <https://hyscaler.com/insights/google-musicfx-impact-music-creator/>
- Johnson, D., Horton, E., Mulcahy, R., & Foth, M. (2019). Gamification and serious games within the domain of domestic energy consumption: A systematic review. *Renewable and Sustainable Energy Reviews*, 73, 249-264.
- Jouët, J. (2018). Digital feminism: Questioning the renewal of activism. *Journal of Research in Gender Studies*, 8(1), 133-157. doi:10.22381/JRGS8120187
- Kember, S. (2003). *Cyberfeminism and artificial life*. Routledge.
- Lorde, A. (1984). *Sister Outsider: Essays and Speeches*. Crossing Press.
- Lyon, D. (2009). Surveillance, power and everyday life. In R. Mansell (Ed.), *The Oxford handbook of information and communication technologies* (pp. 449-472). Oxford University Press.
- Manyika, J. (2022). Getting AI Right: Introductory Notes on AI & Society. *Daedalus*, 151(2), 5–27. <https://www.jstor.org/stable/48662023>

- Noble, S. U., & Tynes, B. (Eds.). (2016). *The intersectional internet: Race, sex, class, and culture online*. Peter Lang.
- Segrave, M., & Vitis, L. (Eds.). (2017). *Gender, technology, and violence*. Routledge.
- Seneca Falls Convention. (1848). *Declaration of sentiments and resolutions*.
- Smith, E. E., & Kosslyn, S. M. (2007). *Cognitive psychology: Mind and brain*. Pearson/Prentice Hall.
- Solnit, R. (2014). *Men explain things to me*. Haymarket Books.
- Wachter-Boettcher, S. (2017). *Technically wrong: Sexist apps, biased algorithms, and other threats of toxic tech*. W.W. Norton & Company.
- Wade, A. (2023). How it's made: TextFX is a suite of AI tools made in collaboration with Lupe Fiasco. *Google Developers Blog*. Retrieved from: <https://developers.googleblog.com>
- Wajcman, J. (2010). Feminist theories of technology. *Cambridge Journal of Economics*, 34(1), 143-152. <https://doi.org/10.1093/cje/ben057>
- Wiggins, G. A. (2012). Computational creativity: The final frontier? *Frontiers in Artificial Intelligence and Applications*, 242, 21-26. <https://doi.org/10.3233/978-1-61499-098-7-21>
- Wollstonecraft, M. (1792). *A vindication of the rights of woman: With strictures on political and moral subjects*. William Gibbons.
- Zhang, Y. (2024). Utilizing Computational Music Analysis And AI For Enhanced Music Composition: Exploring Pre- And Post-Analysis. *Educational Administration: Theory and Practice*, 30(5), 269–282. <https://kuey.net/index.php/kuey/article/view/2841/1783>

Contributors

Kirkwood Adams is a Lecturer at Columbia University where he teaches the first-year writing seminar. He is the co-translator of Johannes Heldén's *Astroecology*, a hybrid book of poetry and images. Together with Maria Baker, he researches genAI's composition processes. Adams and Baker recently co-taught the writing module of Columbia's inaugural interdisciplinary course, "AI in Context."

Maria Baker is an Adjunct Assistant Professor at Pratt Institute, where she teaches Pratt Integrative Courses focusing on multimodal composition, creative writing and thinking strategies. At Columbia University, she is a consultant in the Writing Center and occasionally teaches University Writing. Together with Kirkwood Adams, she researches genAI's composition processes. Adams and Baker recently co-taught the writing module of Columbia's inaugural interdisciplinary course, "AI in Context."

Dr. Adrienne Carthon-Lyons - A native of Savannah, GA, Dr. Adrienne Carthon-Lyons has built a career as a fair and well-versed instructor of English. Her teaching extends from composition to African-American, American, and Caribbean literatures to issues related to gender/sexuality and feminist studies. She has taught on every level from K-12 to graduate school and has a passion for teaching early career students.

Kathryn (Katie) Conrad is a professor of English at the University of Kansas. She is the author of "A Blueprint for an AI Bill of Rights for Education" (*Critical AI*) and, with Lauren M. E. Goodlad, "Teaching Critical AI Literacies" (*Policy Insights: AI and Digital Inequities, NORRAG*). She has published widely on technology and culture; runs an occasional Substack blog, Pandora's Bot; and contributes images to Better Images of AI. She serves as Associate Editor on *Education Policy for the journal Critical AI* (Duke University Press) and on the founding advisory board for Harvard's AI Pedagogy Project. She is co-director (with Sean Kamperman) of the AI & Digital Literacy project, in partnership with the National Humanities Center. She has spoken on critical AI literacy, AI and writing, AI ethics, and AI risk and compliance and has been a guest on several webinars, panels, and podcasts. In June 2024, she served on an experts' panel advising the UN Special Rapporteur for Education on human rights, AI and education.

Anuj Gupta is an incoming assistant professor in Rhetoric and Composition at the University of South Florida's Department of English, where he researches, teaches, and designs digital writing technologies to make tools like generative AI more usable and human-centered. His contributions have earned him the CWPA's Kenneth Bruffee Award, AAC&U's K. Patricia Cross Future Leaders Award, and Kairos' Graduate Student Research Award. Anuj's work has been published in *Computers and Composition*, *IEEE Proceedings*, *Composition Forum*, *Composition Studies*, and the *WPA journal* (among others) and he is currently co-editing *Digital Literacies for Human Connection*, a forthcoming essay collection with the National Council of Teachers of English. Previously, he helped build one of India's first college-level writing programs at Ashoka University.

Sean Kamperman is an Assistant Professor of English at the University of Kansas specializing in professional and technical writing, disability studies, and digital rhetorics. His research and teaching examine the role of genre and other professional norms, including norms around writing

technologies, in constructing intellectual disability. His current book project, *Genrefying Access: An Infrastructural Study of Access, Writing, and Disability*, uses the lens of genre studies to explore how professional literacies constrain opportunities for meaningful workplace access.

Dr. Kathleen Kennedy is a Professor of Practice, Retailing and Consumer Science at the University of Arizona's Norton School of Human Ecology. Dr. Kennedy's research focuses on human-AI interaction, influence, and cocreation. An advocate of universal design (UD) principles, she is committed to increasing student success through instructor-directed AI personalization, adaptive learning, and experiential learning. Since 2021, she has directed the multidisciplinary PALS (Personalized Adaptive Learning System) research initiative, which involves over 21 instructors and 1,390 undergraduate students. In 2023, she initiated the AI and Teaching and Learning faculty learning community and workshops, and has worked with over 1,200 teaching professionals on integrating AI into instruction.

Stephen Monroe is Chair and Associate Professor in the Department of Writing and Rhetoric at the University of Mississippi. He is an affiliated faculty member in the Center for the Study of Southern Culture and a steering committee member at the Sarah Isom Center for Women and Gender Studies. Monroe co-chairs the DWR's AI Working Group, which plans and pilots new pedagogies. He is a traditional teacher who embraces technology. His previous work in open educational resources was funded by the Hewlett Foundation and in adaptive courseware by the APLU and Bill and Melinda Gates Foundation. Monroe also serves as Director of the Willie Morris Awards for Southern Writing. His book, *Heritage and Hate: Old South Words and Symbols at Southern Universities*, was published in 2021.

Amy Walter is a Research Scholar at the Friday Institute of Educational Innovation at North Carolina State University. Before joining NC State, Amy was a middle school teacher, instructional facilitator, and intervention coordinator. Amy is now researching older elementary and middle school students' foundational literacy and reading comprehension needs; she also works on how generative AI can be used to improve teacher planning and practice.

Marc Watkins is an Assistant Director of Academic Innovation, Lecturer of Writing and Rhetoric, and Director of the AI Institute for Teachers of Writing at the University of Mississippi. He co-chairs the AI Working Group within his department and serves as a liaison with other departments on campus, exploring generative AI's impact on teaching and learning. He blogs about AI and education at *Rhetorica*.

Mila Zhu is Associate Professor of Curriculum and Instruction and Coordinator of the EDUC Program at Southeastern Oklahoma State University. Her current research focuses on ludic scholarship, integrating games and pedagogy to reimagine the role of play in education.