RAPID: DRL-AI: Constructing Understandings of Generative AI and Machine Learning with High School Youth

PI: Antero Garcia, Stanford University
Co-PI: Matthew Berland, University of Wisconsin–Madison
Co-PI: Nicole Mirra, Rutgers University
Doctoral Candidate: Daniel P. Moore, Stanford University

Introduction

Since the wildfire-like uptake of ChatGPT (and other generative AI tools) in recent years, K-12 schools have been short on answers about what to do. Now that a simple string of inputs can yield an adequate persuasive essay, functional code, or personal response in Spanish or French, nearly every subject in schools is facing difficult questions about how to create policy and curriculum related to AI tools. In order for schools to best respond to these rapidly shifting learning contexts, educators and researchers need to understand how young people use AI and learn with and use AI tools. It is from this urgent need that this project explored youth perspectives and approaches to working with and learning alongside AI resources in contemporary learning environments.

This study explored how young people communicate with AI tools. It begins an assessment of how learning built on inquiry-driven exploration of AI tools in secondary classrooms might yield promising practices for the future. For example, getting an AI tool to give you the "right" text often requires multiple rounds of revision and clarification. Likewise, if you want ChatGPT to produce an image of your favorite Pokémon character, you will need to effectively persuade this tool to build an image based on your expectations. In short, using AI requires nuanced interaction in order to effectively engage in meaningful learning practices. Before we can fully develop inclusive pedagogies around AI tools, we need to understand how young people learn to communicate with these tools in the first place and how to bridge the gap before it ossifies into a tool only for the privileged.

Much of the education and learning sciences literature about GenAI focuses on either bias from and within current tools or their use in curricula (Higgs & Stornaiuolo, 2024; McBride et al, 2024; Nash, 2024). However, this research is expanding to consider how malleable the social contexts of youth engagement with these tools might be (Logan, 2024). Recognizing that young people are constantly innovating and surpassing intended uses of varied platforms (Berland & Garcia, 2024), this project brief builds from our exploration of youth civic innovation and expression with a critical lens on GenAI that centers youth over technologies.

Potential Benefit to Society

As noted above, understanding how young people make decisions when using AI is an urgent need, given how accessible AI tools have recently become. The near ubiquity of tools like ChatGPT caught the field of education flat-footed. As districts race to implement AI-related policies and curriculum, educational decision makers are focusing on short term solutions to a long term presence. The set of new tools and platforms through which young people are presently learning and interacting will play substantive roles for much of the foreseeable future.

The majority of interventions being proposed at this moment–from banning GenAI resources to developing curricular models for exploration and support–generally disregard the thinking of the individuals for whom they are intended: the young people themselves. In doing so, districts and educators overlook existing knowledge, practices, and misunderstandings around how AI is being used. This study offers a foundational perspective for AI-related learning and augmentation. Indeed, AI education is a new field of research, and it is changing rapidly. It is in the public eye; teachers and students share both interest and trepidation around it; there are few experts to offer educational guidance. There are, of course, myriad young people inundated with opportunities for using AI, for AI to utilize their work within constantly growing corpuses of training data, and for unsound pedagogical approaches to flourish as a result. This project, then, centers the possibilities of reframing discursive practices and uses of AI by re-prioritizing the youth perspectives about newly emerging technologies. We undertook this project, in part, to provide a baseline set of descriptive and explicit models as well as exemplars for understanding AI education with generative tools.

The longer-term impacts of generative AI across schools are not yet understood. Particularly considering that not all schools or districts fund computer science teachers and education, a need for baseline understanding of student AI use is necessary for guiding content across all grades and subject areas. Likewise, given that AI has demonstrated utility for use in myriad K-12 subject areas, AI is both a prime opportunity for cross-disciplinary learning and an area from which CS education can act as a guiding beacon from which schools might center new learning practices. Currently, there are bills coming up in state legislatures that will require CS and AI education without a full picture of how youth currently use AI or its broader schoolwide implications. Finding and exploring how teachers and students use, define, and reject generative AI will be crucial to expanding the space of possible CS K-12 teachers.

Participants & Data Collection

This first round of this social design-based experiment (Gutiérrez et al., 2020) occurred in a Northeastern US city in Winter and Spring 2024. Participants included five high school students, their teacher, and three university-based researchers. All students and the teacher identified as Black or mixed ethnicity. Four students identified as girls, one as a boy. Participants ranged from ninth to twelfth graders.

The group engaged in five 90-minute sessions outside school hours in a community learning space. During these sessions, participants interacted with individual ChatGPT accounts based on co-designed topics about GenAI's socio-technical and socio-political elements. Each student had access to a chromebook and a ChatGPT account. Each meeting lasted about ninety minutes, and also included time for socializing, unstructured play, and discussion. Meetings were structured based on student inquiries and the stated preferences for how students said they would like to spend time together.

Data collected and analyzed for this project included observations of the ongoing learning activities, approximately 270 ChatGPT queries, audio recordings of sessions, and transcripts of semi-structured interviews held with participants. Our team transcribed data using otter.ai before uploading to Dedoose for qualitative analysis. For this brief, we particularly focused our analysis on student discourse during the sessions.

Data analysis was inductive and iterative, using grounded theory analysis to develop codes and themes (Charmaz & Belgrave, 2012). We undertook a collaborative coding (Smagorinsky, 2008) approach to the data, and our team engaged in thematic analysis of the coding results. From these efforts, three major codes emerged from the data: dissecting GenAI, mimicking GenAI, and resisting GenAI.

Results

Based on our analysis, student writing practices and collaborative thinking flourished by *un*humanizing a digital agent within their learning environment. This occurred across three themes of interrogation: dissecting, mimicking, and resisting GenAI.

Dissecting GenAI

Students demystified GenAI by tinkering with it. They wanted to understand how GenAI works technically. Through the interactions in this setting, these students spent substantial time considering relationships between human-created AI systems and their unpredictability as well as how LLMs pull from materials created by humans. Through un/structured play, students tested GenAI's limits by:

- actively trying to encounter "guardrails," workarounds, and tricks;
- experimenting with un/useful prompts or fact-checking ChatGPT's responses;
- exploring when the model defaulted to images of white people; and
- testing for guardrails specific to topics of race.

These various activities prompted discussion of the "biases baked into" ChatGPT. This included topics such as racism that students said they have personally experienced in interactions on online platforms such as Reddit, the text of which is potentially part of ChatGPT's underlying training model. These activities fostered critical, civically-minded conversations. They prompted students to ask nuanced questions that build from a critical awareness such as "who wins and who loses from this new tool's existence?"

Mimicking GenAI

After this initial inquiry, students' playful engagement with GenAI, found them mimicking and mocking GenAI's formulaic language. Recognizing the rote practices of offering competing perspectives to many of their inquiries, the students started to discuss and then imitate what they perceived as ChatGPT's "bothsidesism" (e.g. "the pros are…. the cons are… the choice depends… it's up to you"). For the students, this lack of a stance by the tool prompted a shared sense of humor; students laughed at ChatGPT's "spineless," noncommittal responses to civics prompts. With "decision-making" left to the students, our analysis illuminated how youth made agentic moves like "turning the tables," requesting that ChatGPT ask them questions. Mimicking ChatGPT helped students identify ways to use it more agentically. They leveraged this agency to use ChatGPT as a tool for thinking through civic topics, such as their city's future as well as the upcoming 2024 U.S. presidential election.

As one activity trying to capture the sentiment of students exploring and interacting with these GenAI tools, students utilized DALL-E to create fake headlines about a futuristic AI takeover. Resulting images (e.g. figure 1) were filled with nonsense words and illogical images. The strange, alien text provoked laughter and delight.

Figure 1: A ChatGPT-generated image from the student prompt "draw a news line that tells us that AI is taking over our world." Ironically, the tool produced a vision replete with typos and nonsense.

Similarly, the students undertook queries to depict their current city in the future (figure 2). These were then followed by extension-focused queries to make them look more "Afrofuturistic" (figure 3). The students discussed the kinds of nonsubstantive and questionable changes between these images, particularly noting the vague murals in the supposedly Afrofuturistic building details and physical attributes of the people within the generated images.

Figure 2: A ChatGPT image from a group-generated prompt in which students requested an image of the future of their city

Figure 3: A ChatGPT image based on students prompting the system to make the image in Figure 2 “more Afrofuturistic.”

Resisting Personified GenAI

Finally, by building on the iterative nature of dissecting and then mimicking AI, the students in this study found utility in the tool by *unhumanizing* it. During a share-out of takeaways in the final session of this first round of this project, all students discussed ChatGPT as "non-human" or "unhuman." The students invoked and built upon the use of terms like “dumb tool” or “robot." One student said, “It’s a computer, it doesn’t work how you work.” Another said, “the things that can be done with AI, humans need to look over it” and “we need efficiency… but I want humans at the center of things.”

This *unhumanizing* aspect of the student work revealed necessary ways youth were able to resist the designed, humanistic features and social relationships built into GenAI designs.From linguistic ticks to human-voiced audio to built in “memory” for understanding participants’ query histories, students recognized that AI tools like ChatGPT, Siri, and Alexa, are built to guide individuals towards feelings of comfort and trust. The activities in this project ruptured this process, recentering feelings of human ingenuity within the young people participating. By situating Gen AI tools as precisely that–*tools*–these young people calibrated the ability to create and generate new knowledge as human-centered, with *un*human resources to occasionally aid this process.

Future Data

Based on our findings, we are revising the activities for our next phase of data collection with a new cohort of students in early 2025. This design work intentionally builds on the preliminary insights from this first cohort of young people. We are particularly focused on the ways they innovated “everyday” uses of GenAI tools within lived, cultural contexts (e.g. Gutierrez et al., 2017).

In order to diversify and broaden the insights from this project, we intend to conduct this work at a different location in the U.S., exploring how new youth approach AI and how a months-long difference in cultural attitudes toward AI remediate this learning experience.

Implications for the U.S.

This study contributes to a baseline understanding of how AI Education might engage with civics and critical AI studies. Students came into this project with pre existing perspectives about AI that have largely been overlooked by most educational decision makers in this country. By centering their beliefs and providing a context for guided exploration and learning "under the hood" of AI tools, this project demystified these "charismatic" tools (Ames, 2019).

By unhumanizing GenAI, students centered its tools for humanistic purposes. Moving from a civic, dystopic dread, they saw possibilities for GenAI in a civic future that centers humans, with 'robots-in-the-loop'. This paper suggests that a humanizing lens for teaching and learning shouldn't necessarily focus on personifying nonhuman objects. Instead, we can provide students with opportunities to unhumanize GenAI–to mimic it and resist its "humanness." In doing so, these students are realizing a techno-optimism that centers their own creativity, agency, and humanity. With all the talk and fear about the harm that AI can bring, this study spotlights young people who are leveraging GenAI for humanistic purposes: centering youth civic innovation and expression, as well as their agency and humanity.

References

Ames, M. G. (2019). *The charisma machine: The life, death, and legacy of One Laptop per Child*. The MIT Press.

Berland, M., & Garcia, A. (2024). *The Left Hand of Data: Designing Education Data for Justice*. MIT Press.

Charmaz, K., & Belgrave, L. (2012). Qualitative interviewing and grounded theory analysis. The SAGE handbook of interview research: The complexity of the craft, 2, 347-365.

Gutiérrez, K.D., Cortes, K., Cortez, A., DiGiacomo, D., Higgs, J., Johnson, P., ... Vakil, S. (2017). Replacing representation with imagination: Finding ingenuity in everyday practices. *Review of Research in Education, 41*(1), 30–60. https://doi.org/10.3102/0091732X16687523

Gutiérrez, K. D., Jurow, S. A., & Vakil, S. (2020). Social design-based experiments: A utopian methodology for understanding new possibilities for learning. In N. A. Nasir, C. Lee, R. Pea, & M. McKinney de Royston (Eds.), *Handbook of the Cultural Foundations of Learning* (pp. 330–347). Routledge.

Higgs, J. M., & Stornaiuolo, A. (2024). Being human in the age of generative ai: Young people's ethical concerns about writing and living with machines. *Reading Research Quarterly.*

Logan, C. (2024). Learning About and Against Generative AI Through Mapping Generative AI's Ecologies and Developing a Luddite Praxis. In Lindgren, R., Asino, T. I., Kyza, E. A., Looi, C. K., Keifert, D. T., & Suárez, E. (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences - ICLS 2024* (pp. 362-369). International Society of the Learning Sciences.

McBride, C., Lee, C. H., & Soep, E. (2024). "Gotta love some human connection": Humanizing data expression in an age of ai. *Reading Research Quarterly.*

Nash, B. L. (2024). Love and learning in the age of algorithms: How intimate relationships with artificial intelligence may shape epistemology, sociality, and linguistic justice.

Smagorinsky, P. (2008). The method section as conceptual epicenter in constructing social science research reports. *Written Communication*, 25(3), 389–411.