Critical Posthumanist Literacies in the Age of Complex Computation: Reading, Writing, and Living Ethically with Everyday Artificial Intelligence

By

Sarah K. Burriss

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Learning, Teaching, and Diversity

May 12, 2023

Nashville, Tennessee

Approved:

Kevin Leander Ph.D.

Corey Brady, Ph.D.

Melanie Hundley, Ph.D.

Haerin Shin, Ph.D.

ACKNOWLEDGEMENTS

This dissertation made it through literal and figurative storms and plagues—COVID-19, new and early parenthood, and many inclement weather days. Though it feels like a miracle that it exists, it is not one; I was able to research and write thanks to the work and support of my family, my friends, my advisor, my committee, my children's teachers, and the students and teachers I worked with.

I would like to thank Jon for being a super-dad, for his unwavering faith in me, and for his jokes. I would like to thank my children, Zira and Hazel, for their love, and their grandparents, Leah, Phil, Charlie, and Andy, for swooping in when we needed them. I would like to thank my parents and my brothers for their unconditional love.

I would like to thank Kevin Leander, adviser extraordinaire, for his support of my intellectual and academic growth, all while never losing sight of me as a person. I would like to thank my wonderful committee, Corey Brady, Melanie Hundley, and Haerin Shin, for supporting my scholarship and being fantastic thinking partners.

Thank you to the amazingly loving teachers who cared for our children while I was working, and to all of the friends who treated our kids like their own family (Chorong & Sangdon, Nadav & Avital, Holland, ZB, Sara & Denard, I'm looking at you). Thank you to my Leander Lab siblings, Laura, Michael, and David. Thank you to my cohort for your love and friendship, and the other DTL parents who gave me comfort, encouragement, and help during all kinds of labor.

I would like to thank the administrators and teachers at the middle school where I collected data for being so generous with their time and trusting me with their students, and the students for being willing to learn with me, and reminding me why I love working with middle schoolers so very much.

LIST OF TABLES

LIST OF FIGURES

Table of Contents

Chapter 1: Introduction

Cynthia's Story

On the morning of Wednesday, June 17, 2015, I attended a meeting in the sunlit first floor conference room in the Main Branch of the Charleston County Public Library. I remember seeing Cynthia, long-time manager of the historic Dart Branch of the library, sitting in a chair by the window, talking and laughing with some of the other branch managers, her good friends from working in the library system for over 30 years. Later that night, Cynthia and eight of her bible study companions were murdered at "Mother Emanuel," a prominent African Methodist Episcopal Church, barely 500 feet away from the sunny conference room where I sat across from Cynthia that morning.

This was not the reason that I left the library two years later to pursue my doctorate, although it is true that public service was starting to feel increasingly dangerous. This tragedy remained lurking but buried—a privilege, I know, to be able to bury it—until I encountered a book that shifted the ground beneath me. I had been interested in digital literacy and how it intersected with social justice, having seen firsthand at the library how access to information and communication via computers and the internet were simultaneously a source of profound joy—hearing the noisy collaboration of pre-teens building worlds together on Minecraft, for example—and a mode of control and oppression—watching the growing line of anxious patrons who needed our computers to submit the online-only applications for subsidized housing, for example. But I did not expect to see Cynthia in this path I had chosen.

I was reading Safiya Umoja Noble's (2018) book, *Algorithms of oppression: How search engines reinforce racism*, when I was jarred by Noble's case example for Chapter 3: "how a mass shooter and avowed White supremacist, Dylann Roof, allegedly used Google Search in the development of his racial attitudes, attitudes that led to his murder of nine African American AME Church members while they worshiped in their South Carolina church" (p. 11). In that

chapter, Noble recounts how Roof claimed that a Google search irrevocably changed him. In his manifesto, Roof wrote that

> 'The event that truly awakened me was the Trayvon Martin case. I kept hearing and seeing his name, and eventually I decided to look him up. I read the Wikipedia article and right away I was unable to understand what the big deal was. It was obvious that Zimmerman was in the right. But more importantly this prompted me to type in the words "black on White crime" into Google, and I have never been the same since that day. The first website I came to was the Council of Conservative Citizens. There were pages upon pages of these brutal black on White murders.' (qtd. in Noble, 2018, p. 111).

Roof had been directed, first, to a White supremacist website. In her analysis, Noble connects the underlying profit motive and the implicit epistemology of page ranking in search results to what happened with Roof: These results, "lay the groundwork...for implicit bias: bias that is buttressed by advertising profits" (116), and they "are not tied to a multiplicity of perspectives, and the epistemology of 'ranking' from one to a million or more sites suggests that what is listed first is likely to be the most credible and trustworthy information available" (p. 118). By seeing the Council of Conservative Citizens front and center, top-ranked, Roof was involved in a complex sociotechnical web that, perhaps, nudged him toward radicalization. Noble concludes that

> A straight line cannot be drawn between search results and murder. But we cannot ignore the ways that a murderer such as Dylann Roof, allegedly in his own words, reported that his racial awareness was cultivated online by searching on a concept or phrase that led him to very narrow, hostile, and racist views. (p. 117)

The collision of my worlds, the shock of learning the connection between algorithms and the pain and loss I saw vibrate through Cynthia's family, her friends, our library system, affected me profoundly. It was not merely the obvious internet access/government policy problem—like

what I had seen on subsidized-housing-application day—that shaped lives, but also the more hidden, subtle forces of an algorithm prioritizing search results. It was a more-than-human problem.

AI Everywhere

I open with this story for a few reasons. First, it was a revelation that shaped my program of research, traceable forward to this dissertation: I had thought of this terrible event as a human tragedy, but I began to see how consideration of nonhuman agents—like search result algorithms—is crucial and consequential for social justice, broadly, and justice-oriented literacy, specifically. Second, it illuminates key ideas that I investigate in the dissertation, like the implications of living with often-biased, largely hidden, and ever-changing sociotechnical systems.

Our everyday lives are entwined with computers and computational systems. In particular, artificial intelligence (AI) is becoming an ever-increasing presence in routine tasks. From performing a Google search to finding directions on our phones, we rely on often-hidden processes of artificial intelligence to guide what we see, where we go, and how we make decisions. As AI and its attendant computational agents and processes—computers, phones, surveillance apparatuses, bots, data, algorithms—have crept into our lives, so too have they become a common topic of debate and imagination in the public sphere and in many academic disciplines.

AI is deeply embedded in our public discourse in several ways. For example, AI is part of the delivery platform for much of our news (e.g., via Google searching and social media platforms), the subject of daily news (e.g., stories about ChatGPT and the ethics of AI-generated art being two recent high-profile cases), and even sometimes the *author* of some news stories (Seabrook, 2019). The popular press has seen an explosion of nonfiction books criticizing artificial intelligence and algorithmic processes (e.g., *Algorithms of oppression* by

Safiya Noble, *Automating inequality* by Virginia Eubanks, and many others). We see AI featured in advertisements online, on our TVs, and in places we traverse and work.

This everyday presence is not a mere curiosity, nor only hype. Everyday AI has consequences—both large and small—for how we read, write, and live. Although I will not address them in depth here, others have written about how AIs are screening job applications and interview videos (Heilweil, 2019) in biased ways, assigning risk scores and sentence recommendations in the justice system (Angwin et al., 2016), determining who gets a loan (Townson, 2020; Miller, 2020), scanning and identifying our faces in public places (Webster, 2021), and more. Of particular relevance to literacy scholars, AIs are involved in our daily writing and reading; AIs perform email filtering and suggest text to finish our sentences and search terms; rank search results; filter and recommend content on social media; grade student work; and serve as writing "tutors." Beyond the classroom, AIs have penetrated professional writing spaces, including and now far beyond Google's Smart Compose, which prompts Gmail users with predictive text (Lambert, 2018). In a 2019 *New Yorker* article, for example, author John Seabrook (2019) wondered about the future of predictive text AIs and tested out a highly sophisticated program to see if its text was up to *New Yorker* snuff (it was not, yet). In only a few short intervening years we have come to the point where (currently) free AI text generators can convincingly write a whole host of text styles and genres, including, to much consternation, an email responding to a recent mass shooting, sent to university students the week I wrote this (Perrotta, 2023).

For most people, AI is a feature of daily life, but many might not know, recognize, or choose it. AI can be difficult to spot, and even harder to understand or interpret. From smart home devices to email filtering, public surveillance to banking, AI is operative in nearly every corner of our lives, but it remains a slippery subject of critique, or even definition.

What is AI?

This question is a central one in this dissertation, and my view is that we are all still, and always, figuring it out and constructing new meanings. Although I cannot provide a settled, simple definition of AI, I can offer ways that various disciplines and scholars have conceptualized it. The term "AI" is both commonly used and contested. Some people take issue with the name, objecting that it is neither "artificial" nor particularly "intelligent." It also has various technological, political, and cultural meanings, depending on whom you ask: is AI a sign of progress, a specific computational technique, a scary murderous robot, an existential threat, a complex sociotechnical system, a news headline, or a marketing tactic? If you consult a tech CEO, a technoskeptic, and a computer scientist, you might get any combination of these answers.

> Lanier & Weyl (2020) have argued that rather than a technology, AI is an ideology: 'AI' is best understood as a political and social ideology rather than as a basket of algorithms. The core of the ideology is that a suite of technologies, designed by a small technical elite, can and should become autonomous from and eventually replace, rather than complement, not just individual humans but much of humanity. (n.p.).

AI is so slippery to define in part because it is an ever-shifting target; "the AI effect" captures the idea that AI might even be defined as that which cannot yet be done by computers ("AI effect," 2020). As AI accomplishes more and more, the goalposts for what is "true AI" or "general AI" shift as well. To avoid the confusion and controversy around "AI," other terms have cropped up to specify the technical capabilities of advanced computational technology without having to grapple with the meaning of "intelligence." One such term is machine learning, which Mittelstadt et al. (2016) define as "the capacity to define or modify decision-making rules autonomously" (p.3). Other related terms with technical meanings include deep learning, neural networks, and supervised and unsupervised learning.

However, “AI” has persisted as a label both in popular discourse and in computer science circles (although people in the latter do tend to be more specific in its application, and some prefer to use other monikers, like machine or deep learning, when they apply). It is used for marketing purposes when companies want to capitalize on its association with high-powered computing, and it is used to instill fear in many a science fiction film or novel (e.g., *The Terminator, Minority Report*, HAL 9000 in *2001: A Space Odyssey*, and others).

This Research Project

In this dissertation, I investigate the relationships between students and artificial intelligence (AI), especially as mediated by designed curriculum and instruction. In particular, I aim to better understand how educators might build on what students know and experience to develop critical awareness of our entanglements with AI, and how educators can open up agentic possibilities for young people in the AI-saturated world. In the literature review in Chapter 2, I will build a case for the urgent need for this work, arguing that AI has become an everyday phenomenon with deep implications for education, but has, historically, seldom been taught in K-12 settings (except in specialized computer science courses/curriculum) and remains understudied in literacies research despite a recent explosion in interest in the subject. Given that AI has become such an integral part of our communication practices, I argue that critical education around its uses and social impact is crucial for students, but that this kind of exposure is largely absent—or, at best, unevenly applied—in schools and as a subject of research. This gap—in both research and educational practice—is what I seek to address in this project. The dissertation focuses on one part of a multi-phase design project, a study where I designed a curriculum and taught a unit on AI and ethics in a 7th and 8th grade science class.

Research Questions

Four concepts that organize this dissertation, and will be a refrain in each section, are ontology, entangled agency, ethics/justice, and pedagogy. Although I have separated them for

the purposes of ordering the dissertation and my own thinking with the data in this project, they are interlinked; how we define and understand the stuff AI is made of is consequential for how we see ourselves—and our actions—entangled with them, and what we think is right or just. Although I will spend the first parts of the dissertation separating these concepts for practical and analytical purposes, I will knit them back together in the final chapter.

The overarching question that animates this project is, What and how should we teach (children and young adults) about/with AI in order to improve human relationships with each other, the Earth, and nonhumans? In particular, I use critical literacies theory and pedagogy, combined with posthumanist theory, to formulate the following research questions for the dissertation study:

1. (Ontology) What might we learn about AI as a complex, everyday phenomenon/object/system through the enactment of this curriculum?
2. (Agency) (How) can this design and my teaching deepen students' understanding of their agency as entangled with AI systems and open up the potential for agentive participation in/resistance of these systems?
3. (Ethics & Justice) How can I design curriculum and teach to foster ethical engagement around AI?
4. (Pedagogy) How can I incorporate science fiction consumption and production, alongside opportunities for technical engagement with AI, to foster agentive participation and ethical engagement around AI?

About Me: Researcher Role & Assumptions

In the dissertation study, I was a curriculum designer, teacher, and researcher. These roles were shaped by my identities and experiences as a White Jewish woman, a mother, and a former public library worker. "Tzedek, tzedek, tirdov" (translation: justice, justice, shall you pursue) is one of the few Hebrew phrases I have in my lexicon. The pursuit of social justice, and

the insistence on “tikkun olam” (repair of the world) is a hallmark of the brand of Judaism I grew up with and keep choosing as an adult. Steeped in lessons about the depths of human cruelty—I suspect the Holocaust is a persistent specter for most Jews—and the command to try to fix it, it is no wonder that I am trying in this dissertation, and with the help of my students, to figure out what is just, who gets to decide, and what can go wrong if we are not vigilant.

I worked in a public library for five years, serving in the young adult and adult service departments (and filling in when I could in the children’s department). This experience is what landed me here, studying literacy practices and working with middle schoolers. My training in anti-censorship and nonjudgmental, confidential information service for anyone who walks in the door has deeply influenced the type of curriculum I design and the kind of educator I am. I have also seen and felt the joy radiating from a child (or adult) when they find the book or other information they are seeking, or when they talk about their favorite thing to read. Contrary to popular belief, too, being a librarian is not all about books and reading. (This, I am happy to report, is true also of literacy studies writ large.) It was also about providing public service—a safe place to be, a connection to the internet, even just a sympathetic ear. Increasingly, too, our work was digital—helping patrons make email accounts and fill out job applications available only online, providing a fun, safe place for kids to play networked computer games, sifting through information online. This mixture—of literacy practice as a public service where relationships are key, of loving and sharing literature, and of seeing the shift toward digital everything—inspired this work.

Perhaps nothing has affected me more than birthing and raising my children. It has shifted my priorities and given me a lens through which to view everything, including, of course, this work. The effects range from how I approach other people’s children (as people, and as students) to how I think with theory. For an example of the latter, I could theorize all day about the idea of the individual as neoliberal subject versus posthumanist mangle but I have had no

better challenge to these ideas than, say, sending my child off on their own to school or signing a daycare form that says I am solely responsible for my child's actions. These conflicting ideas—of your children as your own, as tied to you, and of your children as independent beings—are borne out every single day right before your eyes, challenging the theories you thought you understood about what it means to be a person. Take privacy, too, as a test case. What does privacy mean to the pregnant person? How can you be alone and apart when you are not just you, when your every heartbeat, much less conscious decision, is tied to your embryo/fetus/child? Birthing and caring for children—always and in this particular moment in America—comes with its challenges, joys, and new perspectives. I understand viscerally that each student in my class/care is someone's universe, and also their own. I strive to remember this when I design curriculum and when I teach.

My work is deeply influenced by these identities and experiences in myriad ways, but I would also like to highlight what I am not; I am not a (current or former) classroom teacher, so I relied on the guidance and assistance of professional teachers—to whom I am deeply indebted—in my instruction. There is much to critique in my instruction as a novice teacher, but (mercifully) that is not the primary focus of this dissertation. I do, where possible, try to explain my reasoning for making certain curricular and instructional choices and how these choices played out in the classroom.

Navigating This Dissertation

There are five remaining chapters of this dissertation. In the next, Chapter 2, I will describe and synthesize relevant extant scholarship on AI in literacy and education and outline my theoretical framework. I discuss the recent explosion in interest in and scholarship about teaching and using AI in schools, but the relative lag in literacies work that addresses these topics. In describing my theoretical framework, which I call "critical posthumanist literacy," I focus on theorizing key concepts that echo throughout the dissertation: ontology, agency, ethics

and justice, and pedagogy. These concepts also organize and animate the findings in Chapters 4 and 5.

In Chapter 3, I discuss the methodology of the current study and how it fits into a larger design project for teaching about AI and ethics. I describe the key goals and design features of the dissertation study, a quarter-long unit on AI and ethics for 7th and 8th graders. I collected video, audio, and artifact data over the course of teaching the unit, along with writing field notes reflecting on my teaching. In my analysis, I employed a combination of traditional qualitative methods, like coding, and more poststructural methods, like Jackson & Mazzei's (2012, 2013) thinking with theory. In this chapter, I also address how my theoretical commitments relate to, and sometimes conflict with, my analytical plans, and my attempt to reconcile them.

Chapter 4 is one of three chapters devoted to discussing findings. In Chapter 4, I address two big, and closely related, questions of the project: For these middle school students, what and where is AI? I discuss how the students broadened—perhaps too much—their ideas of what and where AI is, and how participating in activities designed to showcase AI as a technical and speculative phenomenon played out in the classroom.

Chapter 5 continues the reporting of findings, addressing the next big project questions about agency and ethics/justice we explored in the unit: What does AI have to do with me, and what can I do about it? and How should we live with AI? Here I discuss student attitudes about their data and about their sense of power and ability to make change. Moving to ethics and justice, I explore how students applied normative ethical lenses and created their own codes of ethics by way of engaging in ethical thinking practices and understanding major topics in AI ethics.

Chapter 6 is the final findings chapter, where I address the question: How should we teach about AI? In it, I focus on science fiction pedagogy practices I used with these middle grade students, showing three primary functions that engaging in speculative production served:

as a foil for understanding what was currently possible with AI, as a way to engage in ethical speculation/cautioning, and practicing/preparing for the future.

Chapter 7 concludes the dissertation with a discussion of the findings and their implications for future work. I synthesize across the findings chapters, highlighting challenges that traversed them. I then discuss implications for critical posthumanist literacy pedagogy and raise enduring questions that came out of this work. I end by returning to where I began, rethinking the discussion of Google search as a consequential literacy practice through the lens of critical posthumanist literacy.

Chapter 2: Literature Review and Theoretical Framework

In this section, I will synthesize literature on AI systems in education and in literacy. I argue that AI is ethically consequential for all students in schools and for communication more generally, and is not merely a computational technique belonging to computer scientists and in computer science curricula.

In the Introduction, I pointed out the various ways AI has become a ubiquitous and consequential part of our everyday lives. In the following section, I focus specifically on how AI has become an important agent/topic in education, but historically not in ways that allow students to critically examine or have a say in its applications. This is changing rapidly, however, and there has been an explosion of published work on the topic in the last few years.

Reviewing Literature on AI and Education

AI is present in students' lives and in schools. Beyond the examples of email filtering and social media that travel between home, school, and other spaces (as the internet and smartphones allow), there are specific ways that AIs are being deployed in educational contexts, often without student knowledge or consent. In this section, I will describe several of the ways that AIs are impacting students' and teachers' lives in the classroom and beyond. I will begin with ways that AI is present in schools outside of the curriculum—how it is used to surveil, evaluate, predict outcomes for, and assist in teaching—and end with how it has begun appearing in more classrooms as the subject of instruction, especially in computer science curricula.

Data analytics, including prediction of student outcomes, has become big business in education (Williamson, 2016). This operates both at systemic level—where policies are enacted via "data governance" of education—and by potentially foreclosing individual students' possibilities via algorithmic prediction.

AI also wields power in grading and writing assistance. AI is in widespread use for essay grading in standardized tests, and increasingly in formative assessment software for writing (Nichols et al., 2021). Feathers (2019) discusses how bias in AIs' grading algorithms disproportionately punishes African American students. Dixon-Román et al. (2020) argue that AI writing tools like *Essay Helper* serve as racializing forces in ways that we have not properly recognized and addressed. The influences of AI in writing production and analysis have recently become a topic of enormous and urgent interest with the arrival of ChatGPT. ChatGPT, an AI-powered chatbot produced by OpenAI, recently became available for public use (OpenAI, 2023). Since then, educators and administrators have been concerned with the implications for writing instruction and evaluation. Much of this discourse has centered around concerns of student cheating, although some writing educators are enthusiastic about incorporating ChatGPT into writing instruction. As we have seen with a recent instance of a Vanderbilt administrator using ChatGPT to help compose an email sent to students, there are a wide range of both potential applications and concerns around proper use of this new tool (Perrotta, 2023).

Although natural language processing (NLP) AIs are prominent in school applications, AI is also used for enabling surveillance of students and student work. AI and other surveillance technology is used to monitor, police, and reward student behavior. ClassDojo, a platform for tracking, sharing, and incentivizing students' behaviors, is widely used in the United States (Robinson, 2020). There is even AI that can eliminate the need for the teacher to monitor behavior, detecting whether a student is distracted or concentrating, as reported in a *60 Minutes* episode (Whitaker, 2020). Lockdown browsers are in common use, especially since schools went virtual during the pandemic. Some institutions also require tools like Respondus Monitor, where you also must be recorded via webcam while you are taking a test, and the video is made available to your instructor (Respondus, 2013). These tools use AI to flag any behavior deemed "suspicious."

Despite this presence, and even despite the recent spate of academic and popular press publications, it is unclear whether and how AI makes it into the curriculum for most students. UNESCO released a report in 2022 on government-endorsed AI curricula in use around the world, detailing several that are implemented here in the United States. The curricula they surveyed tended to focus much less on—and devote less instructional time—to issues of ethics and social impact. Unsurprisingly, most of these curricula center and are implemented in computer science or related classes, but the report does address attempts to weave AI instruction into other disciplinary areas, including English Language Arts. As of this writing, there are scant scholarly or peer-reviewed examples of research featuring this (Chao et al.'s, 2022 demonstration of ideas for using StoryQ in ELA being a notable exception), but there is promising work that blends machine learning concepts with discussion of diversity, equity, and inclusion in the works (e.g., Ellis et al., 2022). Outside of academe, there are plenty of blogs, podcasts, and news articles offering opinions on and suggestions for teaching with ChatGPT in the English Language Arts Classroom (e.g., Ferlazzo, 2023; Boyd Waters & Myers, 2023).

For now, these efforts are localized, dependent upon having a teacher who is especially interested or being a research site for the few projects featuring this work. This means that very few students see AI in their ELA curricula. The story is a bit different in computer science curricula and classrooms, although critical uptake of AI's function and applications still seems to be largely dependent on the individual teacher. Articles like this recent report in *The New York Times*, titled "At This School, Computer Science Class Now Includes Critiquing Chatbots," highlight the efforts of select, perhaps exceptional, educators who take it upon themselves to teach critically about/with AI (Singer, 2023). Even here, the teacher they profiled is working against the rules of the school district, which "blocked [ChatGPT] over concerns that it could be used for cheating" (Singer, 2023, n.p.). These teachers take a "critical computing approach," described here as "an analytical approach in which understanding how to critique computer

algorithms is as important as — or more important than — knowing how to program computers" (Singer, 2023, n.p.). But even as we are moving away from focusing solely on programming and toward critique, these efforts remain primarily within computing.

Aside from several major grant-funded research projects aimed at bringing AI into middle grade and high school classrooms (e.g., at UGA and MIT, both NSF-supported projects) (McGinty, 2020; Payne, 2019), and the subsequent publication and use of these curricula internationally (see UNICEF, 2022), a potential exception to this is in computer science curricula. For example, the AP Computer Science Principles course includes a focus on the "impact of computing," which covers some AI ethics (or adjacent) topics (College Board, 2020). This means that an extraordinarily small number of students have any access to meaningful engagement with AI instruction. Particularly if AI instruction is linked to computer science curricula, we must also question how this deepens inequities in access to this instruction.

AI Is an Integral Communication Technology but Is Understudied in Literacies Research

In literacy, we often think about reading, writing, and speaking as forms of (dominantly human) communication. Historically, this has meant that literacy centers human language and the various technologies we use to transmit it. In his work on information theory, Gleick points to how "writing in and of itself had to reshape human consciousness" (p.47) in a shift away from spoken language. Similarly, we can think of AI as a new literacy tool, like writing once was, that will shape our consciousness, and vice versa. In particular, it shapes how we read, write, buy, communicate, decide, treat others, and understand our intelligence and our humanity. However, AI is arguably more agentive—and certainly quicker to act—than a cuneiform tablet. If, as Gleick writes, the work of Aristotle and Socrates to translate ideas into writing "made knowledge more durable stuff than before" (p. 50), AI is making knowledge yet something else altogether. As we saw above, too, there is growing evidence of AI as not just a literacy tool that humans choose to use, but a (co-)author, a reader, an evaluator, and a gatekeeper for which texts we see.

However, there is very little scholarship investigating these interconnections, especially from a literacy standpoint.

AI Literacy Versus AI in Literacy

One great irony of the development of AI work in education is that much work designed to explore the social impact of AI has been called "AI literacy." This is ironic given the absence of AI research in literacy scholarship (although this is rapidly changing). Here I would like to differentiate between "AI literacy" and "AI in literacy," which complicates literature review in this area. "AI literacy" has become a common phrase for being familiar with AI, usually without deep computer science expertise. Long & Magerko (2020) define AI literacy, in proceedings from a computer science conference, as "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" (p. 2). In their 2022 report on K-12 AI curricula, UNESCO describes AI literacy as a combination of two other "literacies," arguing that it "comprises both data literacy, or the ability to understand how AI collects, cleans, manipulates, and analyses data; and algorithm literacy, or the ability to understand how AI algorithms find patterns and connections in the data, which might be used for human-machine interactions" (p.11).

Although these are important concepts for literacy scholars to work with as well, this work does not come from the disciplinary tradition of literacy, rather drawing on the general meaning of "literacy" to mean a mark of familiarity or expertise with a subject. I would like to note here that, while there are key areas of overlap, these definitions of AI literacy do not reliably include core components of how I am conceptualizing critical posthumanist literacy—including speculative thinking, entangled agency, and direct discussion of ethics—discussed in more detail below.

Literacy Scholarship Featuring AI: Writing Instruction, Media Literacy, and Critical Digital Literacies

The surge in attention, and publication around, AI in education broadly and in writing specifically, would be hard to overemphasize. By way of an example, to update my literature review, I searched ERIC, an education-oriented database, for "artificial intelligence." This initial search (for "artificial intelligence" anywhere in the text) yielded over 3,200 results, with more than 700 of those having been published since January 1, 2020 (just over three years). In the past three years alone, there has been an explosion of public interest in and scholarship on artificial intelligence. AI's ubiquity, and, now, popularity in public media and grant-funding opportunities has opened the door for exciting work. However, add "literacy" to the search and these results narrow to just 175 results. Vanishingly few of these results were written by literacy scholars or appear in literacy-oriented journals.

As discussed in the previous section, there is unclear, and definitely uneven access to AI-related curricula, especially in courses traditionally associated with literacy, like English Language Arts. Given the lack of AI-related curriculum, especially in literacy, it is perhaps no surprise that there is little scholarly research that centers it. Within literacy studies, scholars have made notable, but still few, contributions to the discussion of AI. Writing with AI tools, for instance, has been a relatively popular line of study. As far back as 2007, we can find educators in major academic literacy publications talking—albeit uncritically—about the role of assistive writing AIs (Sternberg et al., 2007). Taking a more critical approach, Dixon-Román et al. (2019) view algorithms as "racializing assemblages" in their study of an AI writing tool, Essay Helper (p.1). The authors argue that Essay Helper reinscribes sociopolitical narratives of difference, through its algorithmic construction, its reliance on certain training data, and the use of particular standards and rubrics. In a recent analysis of adaptive writing assessment tool Revision

Assistant, Nichols et al. (2021) caution that these tools must be used thoughtfully so as not to exacerbate inequities in education

Looking beyond composing text, two notable studies examine the relationship of literacy and algorithms (Jones, 2019) and big data (Carrington, 2018). Jones (2019) describes university students' experiences of reading and writing with algorithms and how they understood them as agent, authority, adversary, conversational resource, audience, and oracle. Jones describes how integral algorithms are to modern texts and to constructing readers: "What we read and how we read, and, more importantly, how we are conditioned to imagine ourselves as readers, is increasingly determined by algorithms that operate underneath of the surface of texts" (p. 3). In Jones's view, students needn't understand how algorithms work technically, but rather "reflect upon their own inferential processes when they interact with algorithms" (p.4).

Carrington (2018) uses the case of a young British woman named Sophie to outline new ways to think about critical digital/data literacy in the age of big data and AI. Carrington cautions us about "algorithmic identity" and how algorithmic data categorization inhibits our ability to produce and share our own narratives about ourselves, perhaps leaving a gap between how an algorithm identifies us and how we want to think about ourselves. She draws attention to the ways that the production of data, which is then bought and sold by corporations, is not treated as labor that merits compensation for the producer. Ultimately, Carrington concludes that Sophie is "not able to read the narratives or codes that structure her identity or experience" and she does not understand who is working behind the scenes in constructing her identity, and why (2018, p. 73). For Carrington, being able to identify when, how, and by whom our identity data is collected and used in order to enable personal agency to effect change and avoid exploitation are crucial components of critical data literacy.

Shifting now to work that focuses on developing frameworks and conceptual understandings rather than empirical study, others have adopted data as the subject of their

critical literacies work. Pangrazio & Selwyn (2019) propose another framework, personal data literacies, that draws on prior critical digital literacies work and offers five domains for consideration: (1) Data Identification, (2) Data Understandings, (3) Data Reflexivity, (4) Data Uses, and (5) Data Tactics. This work is important for its serious consideration of data as an object and subject of power in literacies work and how it frames data as dynamic, socially situated, and never neutral. It also provides concrete ways for people to make sense of and reclaim some control of how their data is used. However, the focus on data, like Jones's "algorithms" leaves out important aspects of AI systems. These kinds of targeted frameworks are helpful, especially for developing particular skills relating to personal data, but they are not designed to address some of the larger systemic issues that accompany the interconnections between AIs and humans. We (Leander & Burriss, 2020) offer our response to these demands by incorporating posthumanist ideas of assemblages and networks into our theorizing around machine-human critical literacy practices, using and revising literacy concepts of texts, multimodality, and identity to develop key ideas and new questions for practicing critical posthumanist literacy.

Others emphasize understanding features of the media environment over social criticality. Valtonen et al. (2019) describe "media literacy in the age of machine learning" offering suggestions for some content that that will be important for media literacy education, including content on tracking, recommenders and optimization, dynamic content generation, deep learning, reinforcement learning, attention engineering, and content filtering/curation. Still others advocate for moving beyond "media literacy" to an "ecological approach" (Nichols & LeBlanc, 2021), "situating these phenomena [e.g., fake news, algorithmic bias] as contingent upshots of multi-dimensional, human and nonhuman ecological relationships" (p.13). Drawing on assemblage theory, the authors sketch a dynamic, complex conceptualization of media systems. They use the example of Bratton's (2015) "stack" to guide a sample diagrammatic

effort, demonstrating how the Stack can draw attention to—and thereby open to change—the complex, performative, and overlooked interconnections among media actors. Their civic media ecology is an offering that explicitly moves away from “literacy” (which they link to production, consumption, and mobilization of texts) toward a more sweeping and moving understanding.

There are also digital citizenship curricula that address some related issues, and many freely available lessons online about fake news, deepfakes, and other specific digital media literacy issues. However, it is unclear how, when, where, and who is exposed to these kinds of materials. I was unable to find reliable information on the widespread application and impact of media literacy or digital citizenship curricula (although Common Sense Media claims that its digital citizenship curriculum is used in “70 percent of US schools" on its website, I saw no indication of how they arrived at this number, or what “use” entails), much less their effects on students—qualitatively or quantitatively—and/or information on their adaptability to rapidly evolving technologies like AI (Common Sense, 2023b). Similarly, the National Center for Education Statistics has scant statistical data on these kinds of curricula, with the most relevant being a 2018 study of 8th graders’ “computer and information literacy” skills that reports on standardized testing that evaluates individual students’ performance on a narrow set of computer-based tasks (Fraillon et al., 2020). Moreover, this kind of reporting tends to emphasize individual computational skill to the neglect of a more holistic understanding of students as embedded in sociotechnical landscapes where social factors and agentive machines and materials are at play.

Despite these recent, much-needed efforts, there remains a gap between the realities of AIs and our current educational curricula, practices, and theories. The fact that this critical digital literacies work is in its infancy is both exciting and alarming. Although the scholarship discussed above represents a shift in attention in the literacies field, there remains much work to be done that systematically re-interprets criticality to mean something specific with regard to everyday

complex computing, and applies this to work with students. The work that I have discussed here looks at changes in identity, texts, modality, and critical analysis skills with regard to algorithms, big data, and AI. Very little of this work so far provides us empirical data on how these concepts operate in the world with young people. In addition to more empirical work in this area, given the nascency of literacy work addressing AIs, we need literacies scholarship that not only takes the particulars of life with AIs seriously, but also works to translate those particularities into new ethical understandings.

Theoretical Framework: Posthumanist Critical Literacies

This literacy dissertation is rooted in a critical digital literacy paradigm, while also seeking to reimagine literacy work as a posthumanist endeavor. In the following sections, I outline how I borrow from and blend these two theory bases to create a framework that guides this research.

Before I discuss the theoretical and scholarly traditions that this work is rooted in, I will provide an overview of what "posthumanist critical literacy" means. It is my (though not only mine, see Kuby et al., 2019) supposition that literacy (has and) will become increasingly entangled with machine processes, so much so that it will be difficult to separate out human literacy from machines and their concomitant data apparatuses. This is a fundamentally posthumanist idea, applied to literacy: that we are no longer human, nor necessarily individual, but a tangle of processes involving human and nonhuman matter. I use the term "posthumanist critical literacy" to capture this idea, along with my belief that critical (literacy) orientations will be especially helpful in building the world we would like to live in with these machine processes. (A note: in previous versions of this work, I used the word "digital" along with "critical" and "posthumanist" to describe this line of theorizing, as I work to extend a critical digital literacy tradition. However, I believe that "digital" is no longer a particularly useful distinguishing factor in this work, for three reasons: first, it implies an overly narrow view of posthumanism to focus only

on technology; second, it is applied to such a wide range of technologies that it can be confusing to distinguish what might be included or excluded; and, finally, in its most literal sense, it might even be made obsolete by quantum computing technologies.)

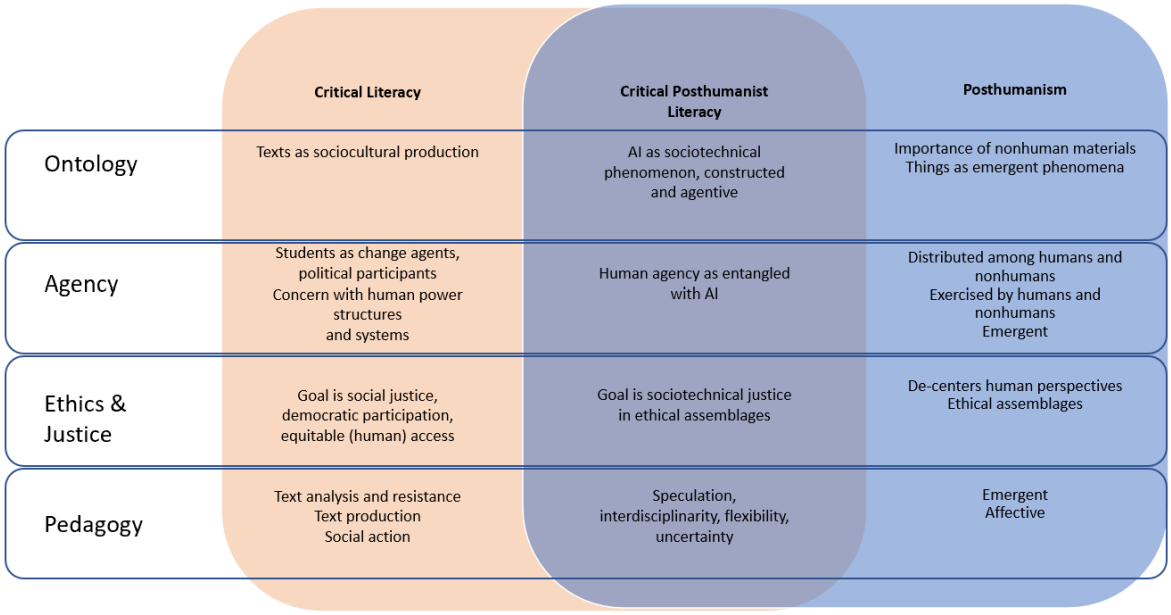
Critical posthumanist literacy, in my formulation, can be summed up by the following set of propositions:

1. AI is a part of posthumanist literacy practice. Our reading, writing, and communication is increasingly bound up with machines and machine processes, including AI.
2. We can use ideas developed in critical literacies to analyze/improve everyday AIs, although no critical literacy model on its own is sufficient to do so. This is because critical literacy models do not sufficiently account for the involvement of agentive machines, nor account for other specific technical capacities of complex computational systems.
3. We can use concepts from posthumanist theorists to extend critical literacy practice into the age of machine writing/reading/living. These include ideas about entanglement in networks/meshes of human and nonhuman materials, and how that can constrain and enable human-machine agency, and understanding social justice as inseparable from technological agents (sociotechnical justice).
4. Critical literacy and posthumanist theories, even when combined, still must be combined with (at least) "just enough" computational/technical know-how to be most useful. For example, knowing how an AI functions allows students to ask more relevant critical posthumanist questions about it. However, what this technical information includes, and how best to teach it, is not settled.
5. Critical literacies have a rich tradition of pedagogical theory and practice that can be helpful in enacting posthumanist ideas. This includes an orientation toward social justice and concern with democratic participation, dialogic practices, critical questioning of power relationships, and techniques for analyzing texts of many types.

This framework arises from a long tradition of critical (digital) literacy scholarship, described below, and seeks to incorporate posthumanist ideas in order to bring critical literacies into the age of near-ubiquitous, consequential AI. I have depicted the framework's main propositions—including where they come from—in the figure below. Alongside these propositions are four core concepts important to theorizing around critical posthumanist literacies that I will define and discuss in the following sections: (1) ontologies of complex computational systems; (2) agency; (3) ethics; and (4) pedagogy.

Figure 1

Blending Critical Literacies and Posthumanist Theories to Make a Critical Posthumanist Literacy Framework



Setting the Stage: Critical (Digital) Literacies

This work is situated within—and seeks to expand—the tradition of critical digital literacy research and practice. From this rich body of scholarship, I draw several guiding ideas for this research, including an emphasis on questioning and dialogue in pedagogy, a social justice orientation, and a focus on digital texts and processes in everyday life. However, I also seek to

extend ideas of critical digital literacy; given the complexity and ubiquity of everyday AI, I argue for incorporating ideas from posthumanist theories of agency into an expanded critical *posthumanist* literacy.

Critical literacy scholarship and praxis has been an influential paradigm for work in literacies. Much of this work can trace its lineage directly back to Freire (1970), whose ideas of liberatory pedagogy centered around education as a way to interrogate and change systems of oppression. In broad strokes, criticality refers to the concern with systems of dominance and power. Importantly, "power" has typically meant influence and control wielded by humans as individuals or in groups. Applying this criticality to literacy, Luke (2012) defines "critical literacy" as the "use of the technologies of print and other media of communication to analyze, critique, and transform the norms, rule systems, and practices governing the social fields of everyday life" (Luke, 2012, p.5). Furthermore, critical literacy is not limited to critique, but has the additional goal of challenging hegemonic forces in the service of changing power structures. Upon reviewing 30 years of research and professional literature, Lewison et al. (2002) synthesized a list of four interrelated dimensions of critical literacy: "(1) disrupting the commonplace, (2) interrogating multiple viewpoints, (3) focusing on sociopolitical issues, and (4) taking action and promoting social justice" (p. 382). These four dimensions also guide this research project, albeit with some additional—posthumanist—specifications.

In digital literacies, these critical ideas have been applied to reading and writing digital texts, and, increasingly, to digital life more broadly. Pandya & Ávila (2013) define critical digital literacies as "those skills and practices that lead to the creation of digital texts that interrogate the world; they also allow and foster the interrogation of digital, multimedia texts" (p.3). They highlight the potential for disruption and challenging dominant narratives via digital media production. Strands of critical digital literacy work have focused on a wide range of topics, including personal data practices (Pangrazio & Selwyn, 2019), social media (Pangrazio, 2013),

civic action (Garcia et al., 2015), digital citizenship (Buchholz et al., 2020), critical media literacy (Kellner & Share, 2005), and the role of platforms (Nichols & LeBlanc, 2020), among others.

Posthumanizing Ontology, Agency, Ethics, & Justice

Human-centered theories have dominated literacy scholarship. This is understandable, as literacy is a human-centered discipline; we study (human) language and its transmission (typically through reading, writing, and speaking)—to other humans. This remains true in this dissertation; I am concerned, ultimately, with human relationships, and I only have access to my human machinery to think and practice. However, this does not mean that posthumanist theories do not apply to literacy scholarship (nor am I alone or first in this effort; see Kuby et al., 2019 for a collection of posthumanist literacy scholarship). I agree with ecologist and posthumanist theorist Morton (2013) when he writes, “Nonhuman beings are responsible for the next moment of human history and thinking. It is not simply that humans became aware of nonhumans, or that they decided to ennoble some of them by granting them a higher status—or cut themselves down by taking away the status of the human” (p. 201). Our continued human existence—writing, reading, and all—depends on nonhumans, so it behooves us to think with and about them.

One of the most difficult and also enticing, for me and perhaps most social scientists, aspects of posthumanist theories like Barad’s (2007) and Morton’s (2013) are rooted in claims about/from physics. For these theorists, postulating about space, time, and quantum phenomena undergirds claims about sociopolitics, and ecological disaster. These theorists use their assertions of physical phenomena to explain sociopolitical phenomena. Having no particular expertise in physics, I cannot justify claiming one interpretation over another (but I do believe that, even for the non-scientists, these can be interesting and perhaps valuable thinking tools/experiments). For example, questions about truth (e.g., How can we tell if a social media post is “true” or “reliable information”?) and reality (e.g., Is this video a deepfake of a politician?

Is this Instagram photo filtered?) are central to certain brands of media literacy. I am open to the idea that physics and philosophy can help us think about/understand/teach about truth and reality, but I do not mobilize these theories to do that in the present work. With that disclaimer, I again would like to borrow ideas from these thinkers even if we do not get to the quantum level. Indeed, this is the strategy that many posthumanist literacy scholars have adopted, eschewing talk of physics and instead applying general posthumanist exhortations like decentering the human, attending to nonhuman agency, and disrupting binaries (e.g., see Kuby et al.'s 2019 volume for this kind of application).

I use these as tools, as thought partners rather than authorities. This does not imply, however, that I use them superficially, skating over ways they do not align. It is the (mis)alignments that I find generative and would like to trouble here. With these caveats, I will borrow heavily from the posthumanist theories of Barad (2007), Bennett (2010), Harman (2018), Morton (2013), and Latour (2005) to lay the foundation for the key concepts in this dissertation: ontology, entangled agency, ethics, and justice. Within each concept section below, I discuss ways that posthumanist theories have shifted humanist understandings of each concept.

Concept 1: Toward Posthumanist AI Ontology

Posthumanist theories give us (at least) two ways of thinking about the nature of things: a process/networked/systems approach and an object-oriented approach. Part of the goal of the analysis in this project is to work out which lens might be most generative—if not "correct"—way to look at AI. Here I will outline the basic tenets of these theories as they apply to AI.

How we conceive of AI, right down to what (kind of a thing or process) it is, is consequential for how we teach and learn about it. Is AI a giant hyperobject, neither reducible to its parts nor categorizable based on its effects, or a complex web of processes, dynamic and flowing? Posthumanist theory gives us ontological grounding for both, discussed below.

Object-Oriented Ontology: AI as Hyperobject

Object-oriented ontology (OOO, or "triple-O") is a poststructural philosophy that has some key departures from the network- or phenomenon-oriented work discussed next: OOO says that objects exist outside of how we perceive them, and outside of their relationships/actions (Harman, 2018). Morton (2013) summarizes OOO in plain, if not totally intelligible, language:

> Objects are what constitute reality. Objects are unique. Objects can't be reduced to smaller objects or dissolved upward into larger ones. Objects are withdrawn from each other and from themselves. Objects are Tardis-like, larger on the inside than they are on the outside. Objects are uncanny. Objects compose an untotalizable nonwhole set that defies holism and reductionism. There is thus no top object that gives all objects value and meaning, and no bottom object to which they can be reduced. (p. 116)

In OOO, objects are those things that are "irreducible in both directions: an object is *more than its pieces and less than its effects*" (Harman, 2018, p. 53).

One useful concept from OOO for thinking of AI ontology is Morton's (2013) hyperobject. A hyperobject is a special kind of object, something so distributed in space and time that one of its hallmarks is being difficult to see from the human perspective. Morton uses global warming as his focal example of a hyperobject, but also explores cities (London in particular) and oil pollution as cases. Ridiculing the idea of hyperobjects as reducible to processes (according to Morton, they only appear this way to lower-dimension beings), Morton offers us the idea of the "mesh":

> A mesh consists of relationships between crisscrossing strands of metal and gaps between the strands. Meshes are potent metaphors for the strange interconnectedness that does not allow for perfect, lossless transmission of information, but is instead full of

> gaps and absences. When an object is born it is instantly enmeshed into a relationship with other objects in the mesh. (p. 83)

Morton focuses on global warming as a hyperobject, but some of his thinking around this example is interesting when applied to AI. Is a drop of rain global warming? No, not by itself, but it is part of the mesh of the larger object. AI is similarly difficult to "see" for how distributed it is across flows of data, human users, and machine processes. So too do human decisions and actions get bound up in this hyperobject mesh. Although Morton talks about hyperobjects as spanning time scales so large as to be incomprehensible to humans, I think that we might extend this also to timescales so short as to be incomprehensible to or impossible for humans, like how AI can process text much more quickly than our human machinery can.

This theory opposes others that conceptualize/ontologize things as being reducible to dynamic connections, flows, or processes. Morton (2013) writes that "A process is a sensual translation, a parody of a higher-dimensional object by a lower-dimensional being" (p. 120), continuing later, "There is a deeper reason why hyperobjects are best seen not as processes, but as real entities in their own right. Seen from a suitably high dimension, a process is just a static object" (Morton, 2013, p. 120). Political scientist and OOO theorist Jane Bennett (2010) offers a view of this concept from another perspective: "'Objects' appear as such because their becoming proceeds at a speed or a level below human discernment" (p.58). She and Morton differ in their view of flow, with Bennett seeing materiality itself as a flow. Bennett argues that everything material, seen properly, is flow, while Morton argues that everything flowy, seen properly, is an object. Morton sees OOO as a "deeper magic": "The magic of systems thinking evaporates in the face of a deeper magic, the magic of real objects that subtend the object system, objects that emit time, space, and causality" (p. 89).

Process-Oriented Ontology: AI as Actor-Network, Phenomenon

The networked view of ontology that Morton (2013) and Harman (2018) write against has a long history in posthumanist theory. The most famous example of this conceptualization is Latour's (2005) actor-networks. Latourian actor-networks emphasize the actor-ness of the network and the network-ness of the actor; they are inseparable, as an actor is also always a network and vice versa. Things are constituted by their dynamic relationships to/with other things, human and nonhuman. In a different but still process-oriented ontological view, Barad (2007) put forward the idea that instead of focusing on things, or representations of things, we should attend more closely to "matters of practicings, doings, and actions" (p.135). This also is an explanation of how things are actually performative, or processes of "enacting boundaries." Barad's conception of agents as formed through a process of "cutting" them differs from OOO's assertion that objects pre-exist any such cuts. In her agential realism, "Matter is neither fixed and given nor the mere end results of different processes. Matter is produced and productive, generated and generative. Matter is agentive, not a fixed essence or property of things" (p. 137). Her "primary ontological units are not 'things' but phenomena—dynamic topological reconfigurings/entanglements/relationalities/(re)articulations of the world" (p. 141).

Using these ideas, we can think of AI as a phenomenon or an actor-network comprised of humans, data, servers, computers, and algorithms all pushing and pulling one another in a constant state of performing AI. This view of ontology, which emphasizes dynamism over thingness, and where we humans can exercise our power by determining where to make our "cuts" to decide what gets considered a specific phenomenon, perhaps allows a little more hope to creep in compared to Mortonian (2013) hyperobjectivity. Rather than thinking of something as so large and impactful as to be hard to grasp, Barad (2007) allows us to consider our human agency in determining what gets to count as part of the object (and in this way, the object is a phenomenon, always being created by cuts).

This line of thinking encourages us to think about how we constitute ourselves alongside our data. The process of making it makes us, and can form a part of the agential unit of "us" depending on where we make our "cut." Can we choose to be separate from our data? How does thinking of ourselves as tangled up in a dynamic network of human-data-AI processes change how we think of ourselves (our identity) and our capacity to make change (our agency)? Most importantly, who or what has the power to enact these boundaries? Latour (2005) can help us think about how we are bound up in and constituted by systems in motion, and Barad (2007) can help us think about how we might "cut" ourselves, or our communities, out—or how others make these determinations for us.

What both of these ontologies do for us is move away from the idea of AI as merely an inert tool for humans to use—it also uses and moves us. By "uses us," I mean that it relies on the data we generate to operate. By "moves us," I mean that it also holds tremendous power to—often invisibly—shape our activities (e.g., of special interest to literacy scholars, our reading and writing).

Concept 2: Toward Posthumanist Agency in Assemblage

To make the case for a more posthumanist conceptualization of agency, I will first describe dominant understandings of agency in education and literacy research, then show how posthumanist ideas about agency allow us to take a new, different look at how we think of student capacities for action in the AI-infused world. I draw on actor-network theory (Latour, 2005) to understand non-human actors as agentive, and on agential realism (Barad, 2007) to understand how identifying relevant, distinct actors is a human process of making "cuts" that opens up certain ethical possibilities and forecloses others.

Humanist Agency in Literacy

The dominant paradigm of agency in literacies work, and in education work writ large, is that of explicit or implied humanism. Furthermore, this humanism often relies on the neoliberal

conception of the individual as a powerful agent (Miller, 2016). This work emphasizes individual responsibility and decision-making, defining agency as an individual's capability or capacity to act. Of course, there are exceptions: for example, a line of research on agency and language learners (Mercer, 2011; Larsen-Freeman, 2019) complicates this narrative by drawing on complexity theory to show how agency is accomplished not by an individual in isolation, nor is a quality that an individual possesses outside of a given context. However, the messaging of individual decision-making and empowerment remains strong in much literacy research.

A dominant paradigm of agency in literacies work arises from the sociocultural theoretical tradition. A key text in this vein (though rooted in anthropological study, it has been taken up in literacy circles) is Holland et al.'s (1998) *Identity and agency in cultural worlds*. Holland et al. (1998) put forward the notion of the "figured world" as a lens for understanding human behavior, identity, and agency. They rely on Inden's (as cited in Holland et al., 1998) definition of human agency as "the realized capacity of people to act upon their world and not only to know about or give personal or intersubjective significance to it," which is "the power of people to act purposively and reflectively, in more or less complex interrelationships with one another" (p. 41-42). In this definition, it is people who act on the world (not vice versa), and there is no consideration of how nonhuman agents might affect human agency. Further, Holland et al. (1998) point to Vygotsky for their understanding of the role of objects, emphasizing the primacy of human interpretation and assignation of value to objects: "Vygostsky … paid special attention to the role of tangible objects, made collectively into artifacts by the attribution of meaning, as tools that people use to affect their own and others' thinking, feeling, and behavior" (p. 50). This account positions objects as "tools that people use," rather than agentive actors themselves.

Similarly, in critical theory**,** technology is primarily treated as a tool or medium we use to achieve or disrupt social dominance, not as agentive in itself. Agency is primarily human, related to a person's standpoint or sociopolitical positioning (Luke, 2012). In critical digital literacies

work, agency has become closely linked to ideas about participation (including “production,” “invention,” and ideas around good citizenship) rather than consumption as a key feature of critical literacy. For example, Mirra et al. (2018) describe how critical digital invention positions “young people as not simply masterful and critical consumers, producers, and distributors of digital literacies, but as inventors with the competencies and dispositions needed to dream up digital forms of expression that adults cannot yet imagine” (p. 17).

“Digital citizenship” has also become a popular framing for work that foregrounds production and participation as an important part of being a good citizen. In a piece responding to the changes in digital life due to the COVID-19 pandemic, Buchholz et al. (2020), position “digital citizenship” as a “move beyond” digital literacy. Their ideal of digital citizenship “requires individuals to confront complex ideas about the enactment of identities and dialogue online as citizens who collectively work for equity and change” (p. 12). Like other current work in media literacy that critiques the “protectionist approach” that seeks to shield children from online harms, they lament the proliferation of “lists of technical proficiencies and simplistic rules for safely engaging online” (p.12). In these formulations, participation with digital literacies is not about learning how to use a technological tool safely, but rather aims toward molding empowered citizens. This is a valuable perspective, especially as protectionist approaches tend to underestimate the agentic potential of young people. When we position young people as important contributors to society, we are not merely recognizing a truth but also expanding the potential for justice-oriented work.

However, this work does not offer us much in the way of understanding digital participation or citizenship given the realities of everyday AIs. The conceptualizations of “digital citizenship” and “critical digital invention” we see here treat technology as a tool for enacting human power, rather than an agentive force in itself. For this reason, I call upon posthumanist

theories of agency to buttress the humanist formulations we have grown so familiar with in critical digital literacies work.

Posthumanist Agency à la Latour and Barad

Posthumanist theories offer an alternative formulation of agency that situates the human among other, nonhuman, actors. One particularly influential theorization of agency, which I touched on in the discussion of posthumanist ontology above, comes from actor-network theory (ANT). At its core, ANT advocates for the return of agency to things as well as people. Latour, one of ANT's founders, defines the nature of action as something "not done under the full control of consciousness; action should rather be felt as a node, a knot, and a conglomerate of many surprising sets of agencies that have to be slowly disentangled" (Latour, 2005, p.44). Thus, action need not be shackled to consciousness, or, by extension, intention. An actor is *any* entity that acts, that does work in/on the world:

> An 'actor' in ANT … implies *no* special motivation of *human individual* actors, nor of humans in general. An actant can literally be anything provided it is granted to be the source of an action...actors are not conceived as fixed entities but as flows, as circulating objects undergoing trials, and their stability, continuity, isotopy has to be obtained by other actions and other trials. (p.373-374)

Even my word "entity" above is misleading and inadequate; the most seemingly discrete and solid "thing" is better captured as a coalescence of associated heterogeneous materials, to be exhaustively and painstakingly traced as networked. Actors are inextricable from the idea of the network in ANT.

Latour (2005) also insists that intentionality is not a prerequisite of agency: "If action is limited a priori to what 'intentional', 'meaningful' humans do, it is hard to see how a hammer, a basket, a door closer, a cat, a rug, a mug, a list, or a tag could act" (p.71). Latour also makes sure to point out that though they are important in acting in the world, objects might not "cause"

or “determine” what will happen. On the other hand, they also shouldn’t be relegated to serving as “mere intermediaries,” but rather “things might authorize, allow, afford, encourage, permit, suggest, influence, block, render possible, and so on” (p.72). This lifting up of the object is not intended to demote the status of the human, but rather to restore the balance from a time in sociology when the material world was largely ignored in favor of the human-semiotic.

Law (1992) also adds an interesting take on “social agents” as always networks and human “agentic” practices (like thinking) are never accomplished by some essentialized individual person/body:

> But it insists that social agents are never located in bodies and bodies alone, but rather that an actor is a patterned network of heterogeneous relations, or an effect produced by such a network. The argument is that thinking, acting, writing, loving, earning—all the attributes that we normally ascribe to human beings, are generated in networks that pass through and ramify both within and beyond the body. Hence the term, actor-network—an actor is also, always, a network. (p.384)

Actor and actor-network are inseparable, as actors never accomplish anything—or have agency—on their own, or outside of emerging entangled human-nonhuman networks.

Unsurprisingly, given both its popularity in general and its origins in technology studies in particular, researchers have applied ANT to examining algorithms, big data systems and, to a lesser extent, analyses of AIs. To illustrate the utility of ANT-like agency to AI-related processes in particular, I will discuss an application where Tufekci (2015) considers an algorithm as an actant which exerts agency. Tufekci (2015) uses ANT to drive her conceptualization of “computational agency” which accounts for how the “labyrinthine components, processes, and opacity, and the subjective nature of the decisions reached” by algorithms and computational agents makes this agency different from other processes (p.207).

Tufekci (2015) is primarily concerned with what she calls “algorithmic gatekeeping,” which is alarming for its invisibility and its ability to tailor results to individuals. For example, a Facebook newsfeed is gatekept in ways that users usually don’t know and in a way that can present differently to each individual user. Tufekci uses two cases to outline her argument for mistrust of computational agency: the role of Facebook algorithms in the Ferguson protests and election tinkering. Tufekci shows that, in both cases, algorithms were deployed in service of invisible gatekeeping. She warns that “[algorithms] are armed with our data, and can even divine private information that we have not disclosed. They are interactive, act with agency in the world, and are often answerable only to the major corporations that own them” (p. 217).

However, even the strictest Latourian (2005) interpretation of agency does not absolve us of considering human influence. In the process of acknowledging machinic processes as agentive, we should not lose sight of the human actors who played and play a role in the creation, maintenance, and interpretations of algorithmically agentic results, even after they are “set free” in the world. Neglecting to account for the interactions of humans and algorithms runs the risk of implying that algorithms somehow arise out of and act within the ether. The fact that a machine spits out a result—and, to be sure, exerts agency—does not absolve humans of responsibility for their roles in the creation, implementation, and interpretations of those results.

Posthumanist Agency à la Bennett

Latourian (2005) formulations of agency have been taken up by other prominent posthumanist theorists. Jane Bennett (2010) draws on Latour to describe her “vital matter,” asking

> How would political responses to public problems change were we to take seriously the vitality of (nonhuman) bodies? By “vitality” I mean the capacity of things—edibles, commodities, storms, metals—not only to impede or block the will or designs of humans

but also to act as quasi agents or forces with trajectories, propensities, or tendencies of their own. (p. viii)

Bennett (2010) describes a "distributive agency," drawing on Deleuze and Guattari's concept of "assemblage." She defines an assemblage as "ad hoc groupings of diverse elements, of vibrant materials of all sorts" that "are living, throbbing, confederations that are able to function despite the persistent presence of energies that confound them from within" (p. 23-24). Bennett uses the electrical grid as an example of an assemblage, describing "a volatile mix of coal, sweat, electromagnetic fields, computer programs, electron streams, profit motives, heat, lifestyles, nuclear fuel, plastic, fantasies of mastery, static, legislation, water, economic theory, wire, and wood—to name just some of the actants" (p. 25). Bennett uses the lens of the assemblage and distributed agency to analyze what happened during a massive blackout in the US in 2003. She highlights the interconnections of human and nonhuman actors and the confluence of factors that led to this unanticipated event, showing how, in her analysis, agency is distributed among them rather than localizable to single objects or people.

Concept 3: Toward Posthumanist Ethics & Sociotechnical Justice

Both critical and posthumanist theories have their strengths and shortcomings. Particularly relating to this dissertation, critical theory offers little in the way of understanding the importance and complexity of the nonhuman world, and posthumanist theory has burgeoning, but not yet fully realized, accounts of how human power operates. These are distinct orientations, and much ink has been spilled arguing for using one or the other. However, both offer useful lenses for examining the interplay of the human and nonhuman reality of our daily lives with AIs. In this section, I argue for incorporating posthumanist ideas of ethics into what has typically been a critically-dominated paradigm.

A general definition of "ethics" is "systematizing, defending, and recommending concepts of right and wrong behavior" (Fieser, n.d.). (Note: Following the lead of renowned ethics scholar

Peter Singer (2011), I use “ethical” and “moral” interchangeably.) Typically, humans are understood to be the ones doing the systematizing *and* the behaving (rightly and wrongly). However, work in both the ethics of AI and in posthumanist scholarship challenges this assumption. Scholars studying the ethics of AI have challenged both human separability from the operation of AIs and also humans’ status as the only ones capable of moral agency (Behdadi & Munthe, 2020). Given these new developments, it makes sense to draw on new ideas of ethics that do not assume human primacy in all cases. In addition to drawing on posthumanist theories that provide a way of thinking about ethics, writ large, as entangled with nonhuman agents, we can also draw on work from the field of AI ethics, discussed below.

Building from Critical Literacy’s Ethics and Drawing on Disciplinary AI Ethics Scholarship

In critical literacies work, we aspire to be ethical, invoking ideals of justice, democracy, and equity frequently. However, we rarely examine the ethicotheoretical basis of these claims and aims (with some notable exceptions around cosmopolitan ethics —see Sabey & Leander, 2020; Hull & Stornaiuolo, 2014), relying instead on normative and/or implicit definitions of these terms. This is a hindrance when we try to handle the massive changes that have accompanied complex computation, and our entanglement within it. For this reason, along with the availability of rich and relevant scholarship in this area, I draw on computer ethics to point the way toward understanding what critical posthumanist literacy ethics might look like.

Just like most industries and academic disciplines, computer ethics, too, has had to reckon with the paradigm shift accompanying the ubiquity of complex computational systems. As Johnstone (2007) writes,

> Standard ethical theory is concerned with the analysis and evaluation of volitional action; with what makes a chosen action right or wrong, good or bad, whether in terms of intentions and obligations as in deontological theory or in terms of outcomes as in

> consequentialism...there are however reasons to doubt whether this kind of analysis based on discrete actions and identifiable agents and outcomes—essentially, the attribution of responsibility—is adequate to address the full range of concerns generated by the many complex contexts in which ICT [information and communication technology] now affects human experience. (p. 74)

Here we see how changes in technological capability have prompted a reconsideration of traditional ethical work in deontology (rules-based ethics) and utilitarianism (consequence-oriented ethics). Both of these ethical systems have typically relied on "discrete actions and identifiable agents and outcomes" (Johnstone, 2007, p. 74), but AIs have complicated these concepts. Of particular concern to computer ethics scholars, as we see here with Johnstone, is how to understand who or what might be deemed responsible given these complications around actions, agents, and outcomes (Johnstone, 2007; Amoore, 2020; Mittelstadt et al., 2016). As ethics scholars grapple with these changes, we can look to them for guidance in our own critical digital literacies practice.

Work in ethics of AI/ethics of algorithms (e.g., Bostrom & Yudkowsky, 2011; IEEE, 2019) will be especially helpful for developing ethical criticality with complex computational systems. Mittelstadt et al. (2016) provide a useful framework for thinking of ethical issues related to algorithms that takes into account their scale, complexity, opacity, and uncertainty, highlighting six main ethical concerns related to the evidence algorithms use and their actions in the world: (1) inconclusive evidence, (2) inscrutable evidence (because so many data points contribute to any conclusion), (3) misguided evidence (where inappropriate conclusions arise from bad evidence/data), (4) unfair outcomes, (5) transformative effects (e.g., personalization algorithms challenging user autonomy by constraining what people see/limiting their choices and changing ideas of privacy), and (6) traceability (especially locating responsibility). Although the authors orient around "algorithms," the framework clearly takes human actors and data (key additional

considerations in my formulation of AI) into account in ways that allow for easy extrapolation to broader application to AI. I use this framework to answer and ask questions in critical digital literacies work in the sections below, as it summarizes key ethical concerns relating to AI's operation and impact.

I draw from these new ethics a concern with the nonhuman as central to human existence—and, by extension, our ideas of justice. In previous work, we (Leander & Burriss, 2020) outline a vision of how posthumanist theory can help us embrace action towards a more expansive notion of social justice via posthumanist ethics:

> Still, criticality must not stop at analysis; it must extend to action. Our vision of posthuman critical literacy has in its sights not just a world where we can point out how computational agents shape us and we them (and even challenge this dichotomy of human/computational agent), but rather a world where human agents can leverage computational machines and processes to become more ethical assemblages with them. Posthuman social justice—where flora and fauna are included as worthy of ethical treatment, as part of our society, and machines and nonhuman materialities are expected to play a crucial role—is the end-game here. (p. 1274)

To think with this idea of ethical posthumanist assemblages, I draw again on Karen Barad (2007), and also on critical literacies work on digital ethics. Freirean (1970) criticality is inextricably bound to anticapitalist ethics, where fighting against oppression, toward liberation and social equality, is the goal of education. Building on this foundation of critical thought, but moving to more current predicaments, Luke (2018) argues for a turn toward "digital ethics." Luke uses a Deweyan definition of ethics as "the codes, norms, and procedures that govern everyday life and interaction, civility, and exchange in institutions, societies, and cultures (Dewey, 2008)" (p. 186), firmly contending that digital ethics must be normative for success. While I agree with Luke (2018) that "the educational challenge raised by digital culture is not

one of skill or technological competence, but one of participation and ethics" (p. 185), I also worry that the reliance on teaching some set of normative ethics removes an important chance for students to engage critically with the ethical mores themselves. One of Luke's (2018) "foundational claims" to set an agenda for digital ethics curricula relies "on a normative model of digital culture," where "ethics cannot exist as a set of norms or procedures for everyday life in digital cultures without a shared normative vision of the good life" (p.194). Luke's vision is one where (presumably, in my reading) educators and curriculum arbiters come to some shared notion of digital ethics which we then pass on, through teaching, to our students. I would like to offer a different option, though—one where the students engage with the negotiation and creation of the ethical norms (meta-ethics), which are not wholly pre-determined by curriculum or educators, instead of only engaging with the outcomes and application of pre-existing digital ethics norms. In particular, I think that Luke does not adequately allow for the complexity of ethics as they relate to human-nonhuman systems, including the (many) areas of uncertainty that a prescriptive, normative, humanist ethics cannot handle. Students must come face to face with this ethical ambiguity, preferably in a safe environment first, in order to develop the kind of criticality that makes room for ethical decision-making in the face of not just the known issues we face today, but those that they will face in the future.

Posthumanizing Ethics with Barad's "Cuts"

By centering normative humanist ethics, Luke (2018) offers a narrow perspective of what digital ethics education might look like. Here, again, I draw on Barad (2007) to help bridge the posthumanist-critical divide, as she offers an account of ethics that takes into account both human power and the complex entanglements with nonhuman others. I am certainly not the first scholar to want the best of both worlds of the critical and posthumanist; Barad herself aimed to draw on "realism"/materialism/physics to make claims about the social world. She avoids the common critique (and perhaps misconception) of ANT that the human gets lost or relegated to

inferior status by attending to the relationships among humans and non-humans and still acknowledging the influence and importance of human discourse. In Barad's formulation of "onto-epistemology," "Objects, entities and phenomena are demonstrably instantiated in and by material practices, produced performatively in concrete situations and thus – crucially – can anchor political actions" (Hollin et al., 2017, p. 924). Here, the nonhuman and human work inseparably to form (or perform) what we consider "political" or "social."

Hollin et al. (2007) point out how Barad's contribution to ethics is often understated. The "cuts" we make in separating bodies and objects have ethical consequences. These "cuts" are not pre-ordained, but rather made interactionally as we act in the world. These agential cuts, or "the separations between words, things, and knowers are real enough but these separations are effects of particular engagements with the world" (Hollin et al., 2017, p. 930). Because we form certain arrangements and ideas of which things are separate, we also foreclose certain possibilities. The cuts can be a tool of exclusion, which is of ethical import. Hollin et al. (2017) explain,

> A focus on agential cuts, therefore, is generative of particular sets of ethical responsibilities; though matter itself has stability, it is still necessary to be accountable for the cuts that created this stability and to grapple not just with the ethical consequences of these cuts, but with the constitutive exclusions that underpin them. (Hollin et al., 2017, p. 933-934)

This reading of Barad's ethics, as stemming from our interpretations and enactments of certain patterns of agential cuts, opens up the possibility of different agential–and ethical—arrangements if we are able to see them as potentially mutable performative patterns rather than ontologically separable or pre-determined distinctions. This re-working of agent into a separable but not pre-figured entity allows us to consider how and why we make the distinctions

we do between tool and person, human and machine, in ways that might help us see different ethical possibilities.

Toward Sociotechnical Justice

I have put ethics and social justice together because they are closely related: ideas about what is just are closely linked to what we believe is right. Critical literacies scholars also often use these terms together (e.g., Shor, 1999; Luke, 2018). Although they are frequently used together and share a core meaning, I have included both for two reasons: first, ethics has a different disciplinary tradition associated with it that is a rich source of materials and tools for curriculum and pedagogy, along with (more recent) connections to posthumanist thinking in posthumanist conceptions of ethics I discussed above; second, "social justice" has a rich history in critical literacies and is due for a posthumanist update.

I offer that we might signal a reorientation toward incorporating nonhumans in our understanding of justice by using the term "sociotechnical justice." This term signals that our ideas about producing equity among human beings in social systems will now always be also inherently technical because of our everyday entanglement with AI. This conception of justice that recognizes we can no longer achieve social justice without technological justice as well. Justice is no longer merely social and no longer exclusively human; it is bound up with ever-changing and growing technologies.

Concept 4: Toward Posthumanist Pedagogy

A major concern of this work is pedagogical: how should we best teach about AI and its impact on our lives? To this end, I provide a review of important concepts in pedagogy that lay the groundwork for my own thinking in this dissertation.

"Pedagogy" has been used and defined in a variety of ways, from a science to an art, and as different from "education" in that the former is organized around certain social goals (Murphy,1996; Hinchliffe, 2001). I am using the term to mean, simply, a theoretically-grounded

understanding of the “how” of teaching. In particular, I focus on trying to design and use a set of theoretically grounded teaching practices that, when mixed with students, content, and activity structures, create meaningful learning experiences.

There is a large body of work discussing “critical pedagogy,” and a relatively smaller, but still significant, one outlining critical *literacy* pedagogy. Although some scholars have cautioned against providing any sort of strict formula for doing critical literacy, there are core practices that traverse this work. Behrman (2006) outlines six types of practices: “reading supplementary texts, reading multiple texts, reading from a resistant perspective, producing countertexts, conducting student-choice research projects, and taking social action” (p.492). Comber (2015) suggests that key “pedagogical moves” in critical literacy are “Repositioning students as researchers of language, respecting student resistance and exploring minority culture constructions of literacy, and problematizing classroom and public texts” (p. 363). There are variations of these sets of practices from different viewpoints, but they share common ideas about positioning students as creators and agents of social change, and consuming, analyzing, and producing texts (broadly defined) in ways that interrogate hegemonic power structures.

Posthumanist Pedagogy

Posthumanist theory offers a different lens for thinking about pedagogy as emergent and affective. Rautio (2019) described emergent pedagogy as centering the unplanned “disturbances” or “distractions” as key sites of learning, rather than pre-designed plans. They emphasize that in this model of emergent pedagogy, learning cannot and should not be “fully controlled,” teaching and learning is socio-material “rather than individual interactions between humans,” and that these ideas necessitate re-thinking individual assessment (p.233). Hickey-Moody (2009) explores interaction with music, literature, and dance as posthuman pedagogy, where “affect in art is a vector of pedagogy” (p. 274). In these two discussions of posthumanist

pedagogy, widening the understanding (and evaluation) of learning from the only human, only individual and exploring affective relationships with art are key practices.

Teaching with Science Fiction

Fiction—and science fiction in particular—has long been acknowledged as an arena for exploring thorny sociotechnical issues, at times reflecting nagging fears of the day, predicting future dilemmas, and sounding dire warnings (Blackford, 2017). In their discussion of speculative fiction pedagogy, Boaz (2020) writes,

> There is a reason that writers are among the first people targeted at the dawn of authoritarianism. Writers—particularly those who work in the genre known as speculative fiction—can offer their audiences inroads to truths inaccessible in other public realms. (p. 242)

This powerful truth-telling function is but one way that encountering speculative fiction can affect us. Some scholars have explored the various ways that engaging with literature in the classroom can shape ethics and morality. Collin (2020) offers four ways: showing students a model of what is right and wrong (which Collin calls “transmitting”), encouraging students to develop their own ethics (“transacting”), allowing students to apply their ethics to various scenarios (“clarifying”), and helping students to understand differences in cultural systems of ethics (“cultural identity and difference”).

In addition to these functions, science fiction can provide a way to exercise human agency without fear of physical consequences. We can “test out” an idea or an AI and see where it takes us without causing offline harm. In physics and philosophy, the idea of a “thought experiment” is not trivial or fanciful, but rather an important mode of research and exploration with key ideas. I suggest here that taking this kind of experimentation seriously, through both consuming and discussing and producing speculative texts, offers young people a chance to engage with highly consequential ideas with relatively low risks to their physical and

psychological safety.

Conclusion

In this chapter, I showed how AI is increasingly present in our daily lives, including in schools. In education, AI is commonly used to surveil, predict, and evaluate student performance. It is increasingly also the subject of curricula, especially in computer science. I argued for more literacy scholarship featuring AI because of its integral role in our current (and future) communication practices. I also discussed my theoretical framework, critical posthumanist literacy, and how I incorporated posthumanist ideas about ontology, agency, ethics and justice, and pedagogy into existing critical literacies conceptualizations of these terms. This framework serves as a guide for the project, from design to analysis to future implications, and I return to it in the final chapter.

Chapter 3: Methodology

Project Overview and Design

In this dissertation, I explore the complex entanglements among students and everyday AIs. Because there simply are not very many opportunities to study this topic naturalistically given the lack of widespread teaching about AI, I developed a design-based study with several phases in middle grade and undergraduate contexts. This dissertation focuses on one phase of the completed project, a unit on AI and Ethics in a 7th and 8th grade class, but the prior and following phases inform the design/redesign of the curriculum and my own understanding of key concepts and practices. I describe the larger project in the sections below.

The Larger Design Project

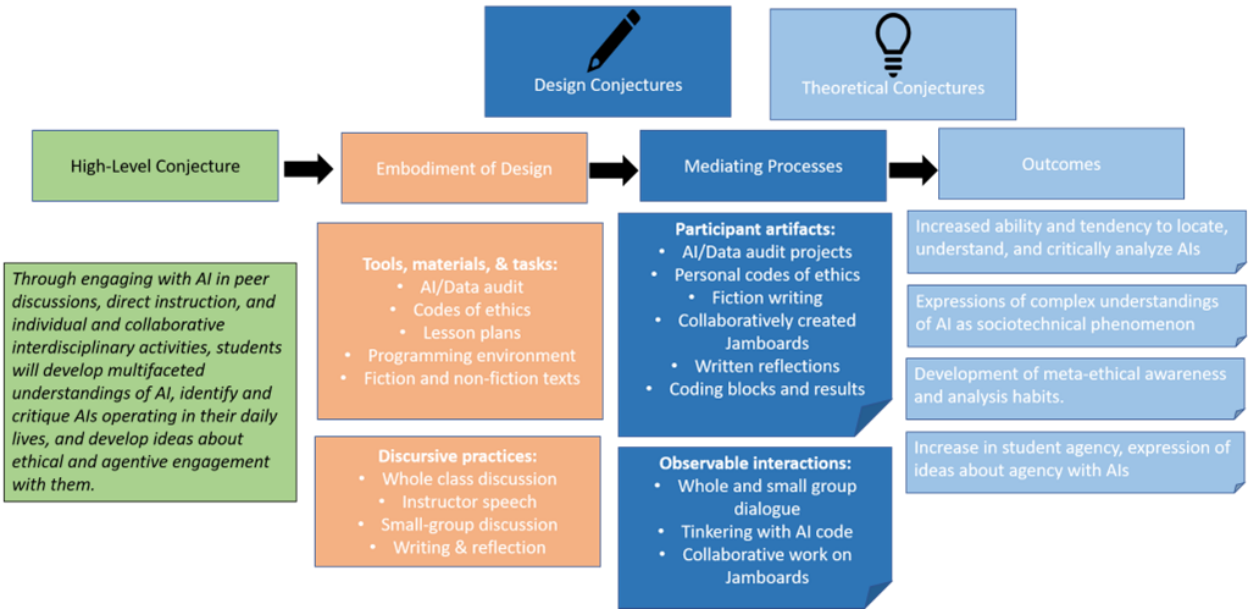
This dissertation began with a design research paradigm. “Design research” can mean a wide variety of things, but here I draw on the common qualities of the work as described by Cobb et al. (2003), where “design experiments are extended (iterative), interventionist (innovative and design-based), and theory-oriented enterprises whose ‘theories’ do real work in practical educational contexts” (p. 13). This paradigm is well-suited for this work because design experiments focus on creating and using materials and activity plans; are “highly interventionist,” where the intent is to investigate the possibilities for educational improvement by bringing about new forms of learning in order to study them” (p. 10); are “test beds for innovation,” which is particularly apt given the newness of my subject (p. 10); and are iterative processes of conjecture, creation, testing, and refinement. This project is also an attempt to meaningfully incorporate critical posthumanist ideas into curriculum and instruction, and to use analysis of this instruction to inform theory-building in critical posthumanist pedagogy.

As part of this process, I created a conjecture map (Sandoval, 2014) and a series of design and theoretical conjectures. These describe the goals of the overarching design project, with the highest-level conjecture being “through engaging with AI in peer discussions, direct

instruction, and individual and collaborative interdisciplinary activities, students will develop multifaceted understandings of AI, identify and critique AIs operating in their daily lives, and develop ideas about ethical and agentive engagement with them." The initial conjecture map is pictured below. For further descriptions of the original conjectures, see Appendix C.

Figure 2

Initial Conjecture Map



The reporting in this dissertation is primarily within phase, reporting on the middle grades AI and ethics unit study, rather than across-phase. Like all teaching (and research), it was messy and incomplete. For example, I had planned to incorporate meta-ethics as a key part of instruction, as reflected in one of my initial design principles (see Appendix C). However, only one group of the middle school students had an opportunity to learn about the term "meta-ethics," as I had a different amount of instructional time with the 7th and 8th graders. For these reasons—reporting on a single phase and shifting design requirements—I have re-worked my initial conjectures into a set of four primary instructional goals, below, that guide and structure the dissertation.

Reframing the Design Conjectures

For the purposes of this report, I have reframed the conjectures, distilling them into four instructional and project goals that both informed curriculum development/delivery and serve as a structure for reporting my findings in the following chapters. In my design and teaching, I hoped to

1. Develop complex, critical understandings of what and where AI is with students;
2. Explore students' personal connections to AI systems, including possibilities for agency and action;
3. Encourage students to think ethically with respect to their uses of and by AI systems and practice applying different ethical lenses;
4. Use consumption, production, and discussion of science fiction/speculation as a pedagogical tool alongside technical activities that demonstrate core aspects of AI's current functionality.

Preliminary Design Phases

This dissertation reports on one phase of a larger design project. The dissertation study sits between pilot studies and a next phase of middle grade teaching set to begin in March of 2023. I describe the pilot phases that preceded the focal study of the dissertation below.

Phase 1: Interdisciplinary Undergraduate Course Study

In the spring and summer of 2019, I conducted an IRB-approved study of a university-wide course exploring ethical issues related to AI. This class was co-taught by two faculty members, from humanities and computer science, and was open to students across the university regardless of year or major. In this phase, I interviewed students and collected course artifacts (discussion board posts). The class met twice a week for 75-minute discussions, had required weekly writing ("blog posts"), technical assignments designed to explore how AI works, and a final project. In addition to observing class sessions, I conducted a total of seven

approximately hour-long interviews, including four with students in the class and three with the instructors and a guest discussant. I also collected and analyzed approximately 60 student blog posts and comments from throughout the semester, and 20 texts written by the instructors, assignments, syllabus, motivation for the course, and other course materials.

These data sources helped me gain an initial understanding of student knowledge and stances toward AI during and after the course. The instructor interviews allowed me to ask about and reflect on effective pedagogy for teaching about AI from an interdisciplinary perspective and centering ethics. In the subsequent year, I served as a teaching assistant for the next iteration of the class, where I gained further experience with the relevant topics, materials, and a new group of students. As I discuss in Chapter 5, I used the organization of curriculum in this course as a tool for thinking about how I wanted to structure my own unit.

Phase 2: Middle Grade Summer Camp Study

During a weeklong camp in the summer of 2021, I served as an instructor for a class on AI and ethics within a camp focused on STEAM topics, specifically "treasure hunting" and preservation. My students included approximately 40 campers going into grades five to eight split into four groups roughly by age. I received camper assent and parent consent from 22 out of 40 campers. The data sources include notes taken by researcher and students' counselors, student artifacts (mostly speculative fiction writing), planning documents, and pre- and post surveys.

This phase was important for developing activities for the middle grade audience, including science fiction writing exercises. It was my first opportunity to work with middle grade students on AI and ethics topics; up to this point, my experience was limited to the undergraduate context. This allowed me to gather information on what the students knew and felt about AI, and also how they responded to activities and teaching materials I borrowed from other curricula and was developing myself and/or with my co-teacher, an experienced

classroom teacher. In working with these students, and especially with the seasoned teacher, I saw the great benefit of well-designed scaffolding materials to guide the writing process, which I subsequently drew from in my writing task designs.

Phase 3: First-Year Undergraduate Seminar Study

I designed and taught a seminar for first-year undergraduate students in the spring semester of 2021. The course was open to students of any major/school. We held synchronous virtual meetings online, eight of which (out of 11 total meetings) were recorded and transcribed. Thirteen out of 20 students consented to participate. Data sources from this phase include class session recordings, course assignments (audit and manifesto), and, for some students, one-on-one follow-up interviews (3) conducted approximately one year after the conclusion of the course. During this phase, I designed and implemented the first round of AI and data audits, which gave me ideas for scaffolding the introduction of this tool for subsequent phases.

Dissertation Study: Middle Grade Unit

These preliminary studies informed the focal phase, where I designed and taught a middle grade AI and ethics unit in an urban, Southeastern private school I have named Overbrook Academy. I taught during the 7th and 8th grade science blocks in the fourth quarter (March-May of 2021) of the 2020-2021 school year. Overbrook is a small, accredited college preparatory middle and high school. Teaching in a private school was not by design, but rather by necessity and opportunity. The restrictions on entering public schools during the COVID-19 pandemic, combined with the curricular flexibility of Overbrook made it a welcome context for this study. Overbrook did not look or feel like a public school in our area, either; from the LGBTQ inclusive posters plastered on the walls to the oddly sized and shaped rooms, repurposed and retrofitted with acoustic paneling and temporary walls, Overbrook was a far cry from the large, cinderblock-and-linoleum public schools common in the surrounding area.

I taught 8th grade on Tuesdays and Thursdays (1:15-3:00pm) most weeks, for a total of 16 visits over 10 weeks, including one visit for one-on-one interviews (six interviews with 8th graders); I taught 7th grade on Wednesdays (1:15-3:00pm) for a total of 10 visits over 10 weeks, including one visit for one-on-one interviews (four interviews with 7th graders). I have included the schedule of sessions, topics, and activities for the 7th and 8th grade classes in Appendix A and B. Out of 13 7th graders, eight participated in study with student assent and parental consent, and out of 19 8th graders, 16 participated in study with student assent and parental consent. All names in this document are pseudonyms, and I have honored student pseudonym requests where possible.

I also conducted a final interview each with two teachers, one who served as the students' science teacher (a kind and caring first-year teacher, who allowed me to use his entire class period), and another who provided support to teachers—and to me, and the students—throughout the school.

Data sources from this study include field notes, planning documents, class recordings, one-on-one interviews with students and teachers conducted at or after the conclusion of the unit, and student artifacts (assignments, answers to writing prompts, samples of notes, survey responses), which I detail in the “Data Sources” section below.

Curriculum Overview and Key Activities

In this section, I provide an overview of the objectives and highlight several key activities in the unit. For a more detailed overview of each day in the 7th and 8th grade units, see Appendix A and B, and for a more detailed description of the teaching activities I feature in the findings sections, see Appendix D. I drew from posthumanist theories and my own disciplinary (literacy, information science) background and perspective to offer an unsettled, multifaceted view of what AI means. I wanted the students to explore various perspectives, introducing AI as

a complex phenomenon that is more than remote computer code, but also a fictional object, a symbol, and an everyday companion in their reading and writing.

Each week had an organizing theme or set of questions, and these differed slightly for the 7th and 8th grade, as I had half the instructional time with the 7th grade class. The overall structure of the class followed the outline in the figure below, but for more detailed information on the objectives and activities for each class session, see Appendix A and B.

Table 1

Overview of AI & Ethics Unit Topics by Week for 8th Grade

Week	Topic Overview
Week 1	What is AI? How does AI work?
Week 2	Where is AI? Where is your data?
Week 3	Where is AI? Where is your data? How does AI classify you and others? What are the connections between “real” and fictional/imagined AIs?
Week 4	What are ethics and ethical lenses? How can we use ethical lenses to think about AI?
Week 5	(How) can we explain how AI works? Digging deeper into ethics: what are meta-ethics? How can we use ethical lenses to think about AI, continued: how can we apply ethical lenses to contemporary AI issues?
Week 6	AI for social good: How can AI be used for education and preservation, and what ethical issues does this entail?
Week 7	Technical perspectives: What is machine learning and how does it work? Science fiction/speculative perspectives: How is AI portrayed in popular media, and (how) is that significant?
Week 8	AI & environmental impact: how can AI harm and/or help the environment? Understanding reports about AI: how can we interpret information about AI when we encounter it?
Week 9	Connecting to current events: How does AI affect our world and how is it reported in the news?

Week 10	Share and celebrate work

To figure out, together, what AI is, I asked students to search for and generate both definitions and examples, and I tried to encourage exploring ambiguity further. I used a combination of direct instruction and exploration (asking the students to find different ways people define AI). We also built computational and fictional AIs of our own. To do this, I incorporated activities including a Jamboard "scavenger hunt" where students collected examples of AI from the news, from a computer science or technical perspective, in science fiction, and from an advertisement or business perspective, then worked with their group to write their own definition of AI on chart paper. I used a slide deck from MIT's DAILy curriculum (MIT Media Lab Personal Robots Group & MIT STEP Lab, 2021) to describe the pieces of AI (from a computational perspective) as input/data, algorithm, and prediction. I also asked students, on different occasions, to "build an AI" two ways: by training a Teachable Machine model to classify objects and by writing a science fiction story.

We also thought together about what data is, where we generate it, and how it might be used. After discussing these ideas together, and brainstorming data-generating and data-collecting activities, I asked the students to complete an "AI & Data Audit." In the audit, I asked students to pick a time frame where they would try to notice and record where they produce data and how it is used. I asked the students to fill out a table where they recorded these encounters with AI and data generation and reflected on what their options might be should they want to change anything about their data and/or AI practices.

I introduced some (dominant Western) ethical lenses to think with, and used materials from the Markkula Center for Applied Ethics (Velasquez et al., 2021) to guide students in thinking through ethical dilemmas from different ethical perspectives. We had whole-group discussions in 7th and 8th grade about a self-driving car dilemma, and a lively discussion on

meta-ethics in 8th grade. I also designed one part (of three) of the 8th graders' final project around thinking of an ethical dilemma relating to their chosen topic/AI application, then working through several of the lenses we discussed and applied in class.

I used and generated a variety of media/materials: slide presentations, video clips, articles, handouts, and a major motion picture. I also borrowed materials from other AI curricula, including MIT's DAILy Curriculum, AI and Ethics for Middle Grades curriculum (MIT Media Lab Personal Robots Group & MIT STEP Lab, 2021; Payne, 2019), consulted with a helpful computer scientist (B. Broll, personal communication, April 18, 2022) and used Google's Teachable Machine. I put all of these materials together in a sequence I designed mostly independently but with periodic consultation with the teaching team at Overbrook. Although I had a plan, I also wanted it to be flexible enough to respond to the needs and interests of the particular students in these classes, the mission and requirements of the school, and what was happening in the world. For example, during the unit an 18-year-old shot 13 people, killing 10, in a Buffalo, NY, grocery store (Franklin & Olson, 2022). He, like Dylann Roof, wrote and published a white supremacist screed online and was active on social media (Yousef, 2022). The next day I taught, I explained the connections I was making between Roof and the Buffalo shooter, and how that linked to the unit. I explained to the students that a major aim of the unit was to give them some practice and ideas for making sense of AI in the news, and asking critical questions about how it impacts our lives and influences how we read, write, and think. After making it clear to the students that they could tap out of this discussion at any time, I provided a guide for small-group discussions at their tables, including questions like, "Who/what is responsible here? What responsibility do social media companies, like Twitch and 4chan have in locating, blocking, reporting this content? What are the risks/downsides of more aggressive content monitoring using AI systems?" and asking them to connect these to the other things we were learning in the unit.

Another mid-unit design I implemented was a three-part project for the 8th grade (see Appendix D for instructional materials related to the project), so that they could pursue a topic of their choosing and, pragmatically, so they had something they could present to their teachers and parents at the end of the semester (a school requirement).

Data Collection

Before each class session, I set up a tablet to audio- and video-record at each group table, and an iPad at the back or front of the room. I used Swivl Teams software to link the main camera and the table cameras. This synchronized the videos so they could be viewed simultaneously on Swivl's platform (see image below). I also used a portable audio recorder as a backup. This setup was wonderful in theory, but had some issues in practice: first, if the wireless connection was broken, the recording would stop and need to be restarted, generating a lot of extra, small video files to sift through and disrupting instruction. Second, the audio quality was poor, especially at the tables. This is probably due to a number of factors, including masks muffling student voices, lower quality microphones on the tablets, and the acoustics of the room. The room was a large room in a historic building that was repurposed as a school. There were two classes that met simultaneously in the room, separated by panel dividers that did not stretch from floor to ceiling. Although the school was taking steps to fix the acoustics (e.g., using absorptive panels, shifting schedules), there were many sessions where the already difficult to transcribe bustle of a middle grade classroom was made more difficult by the setup of the room. I also audio- and/or video-recorded brief final interviews with students who volunteered, and with both teachers I worked with throughout the unit.

In addition to audio and video recording class sessions and interviews, I also collected student work—in paper and digital formats—and photographed all chart paper and white board writing generated during instruction. At the end of the unit, I scanned all of the paper artifacts, then redacted student names from the scanned images.

After teaching sessions, I wrote field notes in a OneNote notebook. These included descriptions of what we did that day, anything interesting or troubling that happened, technical notes (e.g., which tablet stopped working, etc.), ideas for how to adjust instruction, and descriptions of how each day felt to me, what seemed to go right and wrong.

Data Sources

I ended up with hundreds of documents/pieces of data, from pictures of poster paper to interview transcripts, so narrowing down was an important part of the analytical process. Here MAXQDA, the qualitative data management and analysis software I used, proved indispensable and consequential for organizing, sorting, and retrieving relevant pieces of data. This allowed me a flexible and redundant organization system; data could be tagged, labeled, and hierarchically stored in several places at once. I set up case sets for each student, collecting all

of the artifacts/interviews relating to that student together, and I also grouped data by type (interview transcript, class discussion transcript, audit, etc.).

The data sources for the larger design project include surveys, artifacts, and field notes from the summer camp pilot and student artifacts and interviews from the first undergraduate course phase. Although readily at hand in the MAXQDA project file, these were not included in the post-hoc analysis I report on below. This process of bracketing off did not end here; I describe how I sorted and selected data to be considered in this analysis in the following section. The figure below summarizes the data sources that were included for analysis in this dissertation.

Table 2

Overview of Data Sources for the Dissertation

Data Source	Description and Quantity
Products of learning (student work)	Products from in-class activities and assignments, including: ● 20 pre-surveys (answers to questions given at beginning of unit) ● 52 responses to writing prompts given at the beginning and/or end of class ● 15 (pictures of) chart paper from in-class activities ● 14 audit assignments ● 14 movie response sheets (QCQC) ● 12 completed final projects and planning documents from parts 1, 2, 3 ● ~24 science fiction stories and writing planning sheets ● 7 review sheets ● 13 end of unit written responses ● 4 Holocaust survivor AI activity response sheets ● 9 Jamboards ● 23 graph and article interpretation activity handouts ● Miscellaneous notes and other activity sheets

Classroom discourse	Audio and video recordings of class meetings, including: ● Recordings (from up to 5 cameras throughout the room plus additional audio microphone) from 15 visits over 10 weeks in 8th grade ● Recordings (from up to 5 cameras throughout the room plus additional audio microphone) from 9 visits over 9 weeks 5 segments closely transcribed for analysis ● 7th grade self-driving car whole group discussion ● 7th grade fishbowl discussion ● 8th grade meta-ethics discussion ● 8th grade self-driving car whole group discussion ● 8th grade movie response fishbowl discussion ● 8th grade 3 rules for data share-out
Interviews	Semi-structured, with students and instructors, including: ● Brief end-of-unit interviews with 6 8th grade students, ~10 to 20 minutes each ● Brief end-of-unit interviews with 4 7th grade students, ~8-20 minutes each ● End-of-unit interviews with 2 teachers, ~1 hour each
Field notes/memos/planning materials	Including analytical, theoretical, methodological notes Lesson plans and teaching materials

Data Analysis

The most important part of the analytical process was handling the data—watching, transcribing, reading, re-reading, labeling, creating sets—and thinking with it, with theory, and with my (changing) research questions. I touched each artifact, recording, and transcript, obsessed over getting a mumbled piece of speech just right, thought with them, and wrote memos about them. From the fine-grained corrections I made to the transcripts, often listening to/watching a piece of data over and over to get the words just right, or looking at a different camera angle of the same moment to get a different perspective/hear different audio, I got close to the data. I delighted in the funny things the students said and wrote, the doodles they included in their work, and seeing the creative products they made.

In this section, I focus on the post hoc analysis of the data, where I sifted and sorted through artifacts, memos, and transcripts after the data collection phase was completed. However, I also did “microanalyses” and reflection during data collection in each phase and between phases in order to refine and redesign my teaching practices and materials. This analysis happened both in the moment of instruction, as I monitored what seemed to be working or not, and adjusted accordingly, and in later reflection as I wrote field notes, or as I reviewed data from past phases to inform the design of the current phase.

The bulk of systematic data analysis, however, occurred after all data collection for the middle grade unit had been completed. This post hoc analysis proceeded according to a blend of established qualitative methodology procedures, including a process of transcription, organizing, and open coding, and less stepwise “thinking with theory” described in more detail below.

Data Cleaning and Organizing as Analysis

Although often treated as a process before the real analysis can occur, cleaning, sorting, and transcribing data does not precede analysis—it is analytical. As any librarian knows, a mislabeled or misfiled book can be functionally lost forever if the collection is sufficiently large. The same goes for the hundreds of artifacts and pages of transcripts in this project. A good system for classifying and retrieving books and data helped me bring together the wide array of data sources I collected throughout the unit. In this effort, the qualitative data analysis software package MAXQDA was my companion.

MAXQDA allows for coding of image as well as text, so I uploaded scanned images of all artifacts and coded directly on them, as well as transcripts of selected portions of class sessions and interviews. I organized the data files into a hierarchical structure by type/context (e.g., middle grade unit>activity>file name with student pseudonym) and also into sets for each

student so I could easily see all of one student's work together or all students' responses to a particular activity together.

Similarly, transcription is an analytical act. Elinor Ochs (1979) has written about transcription as theoretical work, and she (rightly) acknowledges that "What is on a transcript will influence and constrain what generalizations emerge" (p. 167). What gets included in a transcript constrains your analysis, and the process of making these decisions is consequential. Handling the data by correcting or creating transcripts also brought me closer to the data, and I memoed—writing down things I noticed, interesting patterns and ideas—as I did this painstaking work. I started with a combination of machine-generated (Otter.ai) and human-generated (Rev.com) transcripts, but the poor sound quality and multitude of speakers meant that I had to do a lot of work to correct them and add in sounds and descriptions of pointed looks. For each transcript, I spent hours reviewing video footage and the separate, higher-quality audio recording, sometimes replaying a moment from different camera angles many times to hear a student's contribution. As I could see when viewing my multiple cameras (as in the format above), creating a single written record of what is said necessarily reduces (for good and for bad) the dimensions of the data. In a future project, I hope to incorporate the synchronized video into my analysis.

Data Selection

I was taken by surprise by the analytical import—and the difficulty—of winnowing down my data; I simply had too much. But what if I left out or "lost" an important piece, misfiled as I coded? How could I account for those perspectives that might go missing by not using every medium available (i.e., by using transcripts instead of analyzing video files directly), or those perspectives that were already missing by virtue of how I collected my data and who opted in to the study? Figuring out which types of data needed to be included—and how to express the partiality of the data—became an important step in my analysis. I identified what would become

the “relevant texts” (Saldaña, 2021, p. 36) for each piece of analysis by searching through my codes and re-reading my field notes. I noted relevant/interesting moments in instruction and found those in the video recordings, then sent those segments out for transcription (with my heavy “cleaning” afterward). I also narrowed down the relevant texts by activity; for example, I chose to analyze products from specific activities, knowing I had designed them to address a certain topic/research question. However, I also did not want to limit myself to only those activities explicitly designed to address a certain concept/question, as interesting things happened outside of the design. This is where the open coding of the majority of the corpus became crucial, for not missing out on the undesigned interesting moments.

Coding and Thinking/Writing with Theory

Saldaña (2021) treats coding not as an end in itself, but rather that “coding is in service to thinking.” He defines codes as “symbolic prompts or triggers for deeper meanings that the data may evoke during analytic reflection” (p.107). For me, coding is part mystery (a subconscious meaning-making process), part organizing (labeling and collecting things for looking at them together), and always inventive (adding meaning through researcher-as-research-tool). Some of my coding is simply cataloging, adding metadata to my data to make it easily findable and relevant. Saldaña calls this cataloging “attribute coding,” which “logs essential information about the data and demographic characteristics of the participants for future management and reference” (2021, p.111). Of course, cataloging—as, again, librarians know well—is non-neutral and analytical.

I would call my approach a “semi-grounded” one, in that I used my research questions, experience and lessons learned from earlier project phases, and my theoretical framework to look for certain things in the data, but I also hoped to see new connections and ideas that did not fit within these frames. As I open coded, which I tried to do quickly so as not to overthink it, I revised my research questions and began to formulate ideas about how to keep organizing my

data around patterns and disjunctions I was seeing. This is where I diverged from canon grounded theory or coding practice; I did not continue with axial coding, but rather used a combination of my codes and my data organization structure to select relevant data to read over again and again, honing my research questions and returning to the data and then honing the questions again. During this process, I started writing into my analysis. Some of the magic—it feels like magic to me—of finding the findings happened through this process as well. My analysis did not precede the writing as a finished step, but happened as I wrote, referring to MAXQDA and my coding but also returning at times to the original, raw data to check something out, see the context more clearly.

Close at hand, too, were the texts from theorists I was working to understand and apply, especially Karen Barad's (2007) *Meeting the Universe Halfway*, Morton's (2013) *Hyperobjects*, Latour's (2005) *Reassembling the Social* and Bennett's (2010) *Vibrant Matter*. I was reading and re-reading both my theory and my data as I wrote, in an iterative process of meshing theoretical thinking, data sorting and analysis, and writing. In doing this, I was "thinking with theory" (Jackson & Mazzei, 2012), "plugging in" concepts from theory and data into one another to make something new with/of them both. For example, I read and re-read the student AI & Data Audits, and also my codes of them, with these focal theories in mind, searching for ways students and I might see agency as emergent, distributed among humans and nonhumans, and networked (Barad, 2007; Bennett, 2010, Latour, 2005).

Jackson & Mazzei (2012; 2013) propose this method as an alternative to traditional qualitative processes. They argue that "data interpretation and analysis does not happen via mechanistic coding, reducing data to themes, and writing up transparent narratives that do little to critique the complexities of social life; such simplistic approaches preclude dense and multilayered treatment of data" (p.261). They also emphasize the inconclusiveness of data, describing data as necessarily "partial, incomplete, and always being retold and remembered"

(Jackson & Mazzei, 2013, p. 263). Using this mode of thinking about and with data and theory, I tell a story with my data below, a story that I created with the heavy responsibility of attempting to represent my students, their work, and my own experiences well. It is, as Jackson & Mazzei (2013) put it, "a production of knowledge…emerg[ing] as a creation out of chaos" (p. 263).

Although these techniques—coding and poststructuralist thinking practice—are often described as at odds with one another, I think they can work together nicely. If we treat coding, as Saldaña (2021) puts it, as "in service to thinking," it can be but one other avenue to "thinking with theory" (Jackson & Mazzei, 2013) as long as we remain flexible and open to new categories and interpretations. We can "plug in" our data to the theory, but also our codes to our theory and data, using each idea to think with the other(s). When we see codes as thinking tools rather than reductive categories, they can contribute to more emergent, dynamic analytical processes.

Toward Critical Posthumanist Methods in Design Research

There is an impulse in social science, even among those of us who use interpretive methods, to standardize, demystify, and justify the process of data collection and analysis. I understand this impulse; following steps is reassuring, relying on having an interesting thought is terrifying. There is a historical and political reason for this impulse as well; we have a lot of work to do yet in disentangling ourselves from the research paradigms of biological and quantitative sciences, where the role of the researcher is to remain as neutral and forgettable as possible, so that studies can be reliably reproduced no matter who is doing the work. This project is different; I am tangled throughout the process, from design and teaching to analysis and writing. No one else would have produced these same findings and written about them these same ways, because I am the primary research tool (operating, of course, in a web of materials—like video recorders—and other humans—like the students).

I used a combination of traditional qualitative research moves, like open coding, sorting and organizing using software, and transcribing spoken words into written text, and less concrete and procedural moves like “thinking with theory” (Jackson & Mazzei, 2013). Initially, I had planned a very traditional qualitative analysis, drawing from the canon of grounded theory to envision a stepwise process of open and axial coding, and a neat codebook, complete with inclusion and exclusion criteria, to append, as I had done in previous work. However, as I thought more with posthumanist theory, I began to see tensions between the strictly traditional qualitative methods I had planned and the “posthumanizing” changes I was trying to make in my thinking, teaching, and analysis. I saw posthumanist theories as cautioning against prescriptive moves that might overly simplify, objectify, or close off analysis when things get messy, as they inevitably do in the classroom, and in research. This is where Jackson & Mazzei (2012, 2013) proved very helpful to me in describing the shift I was trying to make away from methods that overly reify and solidify data and do not adequately address the role of researcher thinking; yes, my organizational/coding system was important for helping to put disparate data side by side to think with, but much of the analytical work happened in that chaos that Jackson & Mazzei describe, in the mess of my thinking with theory and reflecting on my experience while handling the data over and over.

I also take seriously my role as not only analyst, but also teacher; I generated data and analyzed it. I shaped my data collection but I am also a part of this data, which I think of with Barad’s (2007) idea of intra-activity; there is no “outside” the phenomenon of this research project, no “objective” perch from which I could observe. This is not a problem or a surprise for the posthumanist, though; teaching and research are always messy, chaotic, sociomaterial tangles, and this was no exception.

Chapter 4: Big Questions about AI: What and Where Is It?

Organization and Overview of the Findings

Chapters 4, 5, and 6 contain the findings of the study. In Chapter 4, I discuss findings related to AI ontology, and specifically how the students responded to activities designed to think about AI as a complex, interdisciplinary, everyday phenomenon. In Chapter 5, I discuss findings related to agency, ethics, and justice. The findings in these chapters are organized around key instructional aims: (1) developing complex, critical understandings of what and where AI is with students; (2) exploring students' personal connections to AI systems, including possibilities for agency and action; and (3) encouraging students to think ethically with respect to their uses of and by AI systems. I have linked each of these three areas to the corresponding theoretical, design, and research elements in the figure below. These categories demarcate porous boundaries; I use them as an organizational structure here, but they are overlapping and inseparable in theory and in practice.

In Chapter 6, I look at the successes and shortcomings of using science fiction production and consumption as a key part of my pedagogy, addressing my fourth project goal: (4) Use consumption, production, and discussion of science fiction/speculation as a pedagogical tool alongside technical activities that demonstrate core aspects of AI's current functionality. Of course, considering pedagogy is a part of each findings section, but I have chosen to make it the star of this chapter; rather than viewing the pedagogical choices I made serving as a way to achieve an instructional end, here I treat a particular pedagogical move—using fiction—as a strategy that travels across the instructional goals, considering links to theory and practice.

Within each of the findings sections, I introduce the relevant research questions and the focal data types, discussing my rationale for designing the curriculum this way. I then identify overall findings from across students and activities. After discussing each of these selected

themes, I link the data and theory, reflecting on my teaching and thinking about how the theory and data, brought together, (Jackson & Mazzei, 2012; 2013) might inform redesign.

Table 3

Mapping Theory, Design, and Research

Mapping Theory, Design, and Research				
	1 AI Ontology Chapter 4	2 Entangled Agency Chapter 5	3 Ethics & Sociotechnical Justice Chapter 5	4 Speculative Pedagogy Chapter 6
Big Project Question	What and where is AI?	What does it have to do with me, and what can I do about it?	How should we live with AI?	How should we teach about/with AI?
Theory to Think With	Posthumanist ontologies: Barad's (2007) "cuts," Latour's (2005) actor-networks, Morton's (2013) hyperobjects, and Bennett's (2010) vibrant matter	Entangled agency: Latour's (2005) actor-networks, Barad's (2007) agential realism, Bennett's (2010) distributive agency	Posthumanist and disciplinary AI ethics, dominant ethical theories, social justice in critical literacy	Science fiction as ethical laboratory Critical pedagogies
Instructional Goal	Students will develop complex, critical understandings of what AI is from a variety of perspectives.	Students will explore personal connections to AI systems, including possibilities for agency and action.	Students will practice thinking ethically with respect to uses of and by AI systems and develop their own ideas about what is ethical with respect to AI.	Instructor(s)and researcher(s) will develop ideas and practices for critical posthumanist AI ethics pedagogy.
Design Elements	Activities: Jamboard AI Scavenger Hunt, Group Definition Activity Direct Instruction: 3 parts of AI slides Dialog: whole and small group discussions	Activities: AI & Data Audit; 3 Rules for Data Activity Direct Instruction: defining data Dialog: whole and small group discussions	Activities: Ethical Dilemma Activity, Project Part 3 Direct Instruction: Ethical lenses/major ethical theories slides Dialog: self-driving car whole group discussion, meta-ethics discussion, chart paper onion "discussion"	Activities:watching, responding to, and discussing science fiction film, writing speculative fiction stories and scenarios in final projects Dialog: Ron's Gone Wrong fishbowl discussion

Primary Research Question	What can we learn about AI as a complex, everyday phenomenon/object/ system through the enactment of this curriculum?	(How) can this design and my teaching deepen students' understanding of their agency as entangled with AI systems and open up the potential for agentive participation in/resistance of these systems?	How can I design curriculum and teach to foster ethical engagement around everyday AIs?	How can I incorporate science fiction consumption and production, alongside opportunities for technical engagement with AI, to foster agentive participation and ethical engagement around AI?
Focal Data	• Scavenger hunt Jamboards • AI definition chart paper • pre-survey responses • 8th grade pre-write, week 1 day 2: What do you need to know to make an AI? • End of unit writing • End of unit interview • Fictional AI inventions	• Completed audits • 8th grade week 2, day 2 session recording • 3 rules for data use (7th and 8th)	• 7th and 8th grade self-driving car discussion transcripts • Ethical decision making handouts • 8th grade final projects • Onion discussion chart paper	• Pre-write: fictional AIs • Speculative fiction writing • 8th grade final projects

What and Where Is AI?: Thinking of AI as "More Than Just Robots"

In this chapter, I discuss findings related to my teaching with activity designs for defining, characterizing, and understanding what and where AI is with students. In particular, the research questions guiding this section are:

- What does AI come to mean for us as a group?
- What can we learn about AI as a complex phenomenon/object/system through the enactment of this curriculum?
- (How) can an audit activity support students in noticing AI in our everyday lives?
- How did student understandings of what and where AI is change throughout the course of the unit?
- Reflecting on instruction and thinking with this data and theory, how might I redesign around the first instructional goal (students will develop complex, critical understandings of what AI is from a variety of perspectives)?

Throughout the AI and Ethics unit, I wanted students to develop an understanding of AI as a complex phenomenon that does not solely belong under the purview of any one discipline—a phenomenon that they, too, can build, think with, and critique. To this end, I designed activities and borrowed from other curricula to encourage students to see how others define AI from different perspectives and to come to an understanding of how we might define AI as a group.

The focal data for this section includes

- the artifacts created from a "scavenger hunt" where I asked them to find examples and definitions of AI from different perspectives and gather them on a Jamboard
- definitions and drawings of AI they subsequently created in their table groups
- their pre-survey responses to questions about what and where AI is
- a pre-writing prompt where I asked them what they'd need to know to make an AI
- products from a writing activity where they designed an AI to put into a story
- end-of-unit interviews and writing responses.

Although the AI and Data Audit is linked closely to looking for AI in students' daily lives, I have left deeper discussion of that activity to the next chapter, where I bring in ideas about agency in relation to the audit process.

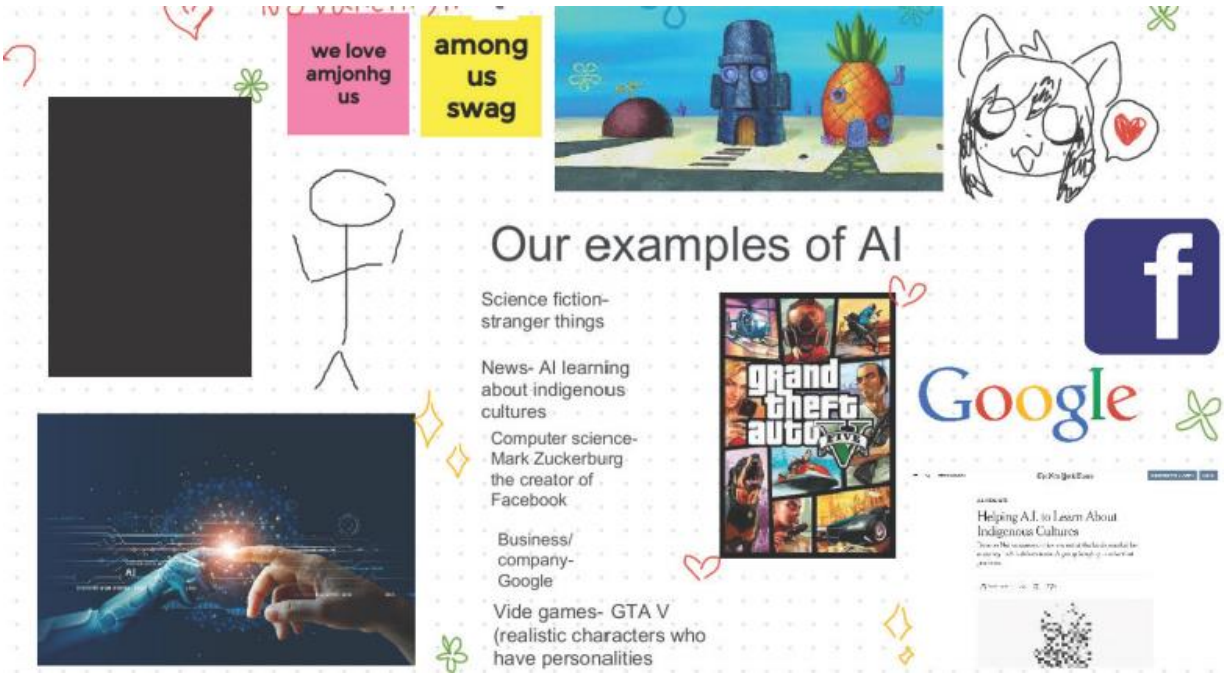
Selected Themes in Defining and Describing AI

As I analyzed this data, I identified three key themes, including a broadening of student views of AI throughout the unit, but a tendency to overgeneralize about where and what AI is; a lingering separation between fictional and technical understandings of AI; and a missing historical or developmental perspective on AI. After I discuss these trends, I reflect on them with my theory and data to offer ideas for redesigning activities that might better work toward the instructional goal of having students develop complex, critical understandings of what AI is from a variety of perspectives.

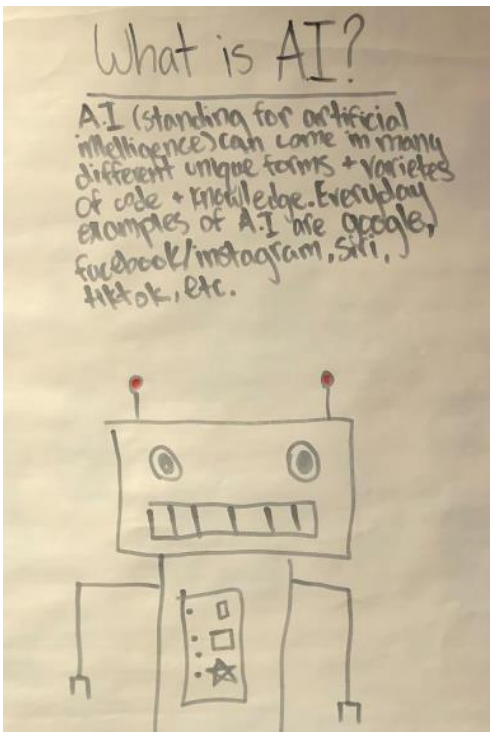
Theme 1. AI as Only Robots or Everything Electronic: From Narrow to Overly Broad Student Descriptions of AI throughout the Unit

At the beginning of the unit, I gave the students an introductory survey to see what they knew about AI and ethics, and what they were curious about. Many students were unfamiliar with definitions of AI, like June, who wondered, “What is AI?” Karl wrote “not sure ive never learned about this stuff.” Other students were familiar with computer programming and everyday instances of AI, like AI personal assistants. Many students associated AI exclusively with robots at the outset.

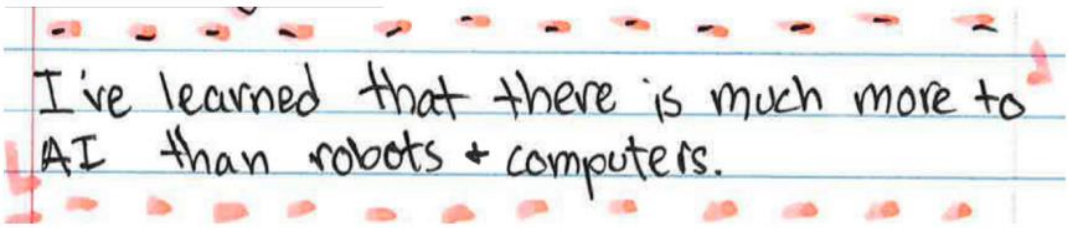
However, this perception of AI as only robots was not long-lasting. On the first day of instruction, I asked the students to gather different definitions of AI from various perspectives (computer science, the news, science fiction, etc.). Students completed a “scavenger hunt” to find definitions and examples of AI from various perspectives. Below is a product of this activity from one group of 8th grade students (Mia, Susan, Jonah, and Trevor). In the Jamboard, you can see a combination of fictional examples from video games and TV shows, corporate/computational examples, and examples of AI from the news. The Jamboard is a multimodal collage of AI representations: pictures, commentary, doodles, student-generated and copied-and-pasted text all work together to paint a picture of AI as a phenomenon with many faces. I highlight this not just because it demonstrates that students saw AI from different perspectives, but also because they used key critical digital literacy practices to do so. In the process of making a multimodal, collaborative text, students not only increased their exposure to different ideas about AI, but they did so by searching for a variety of texts, curating them, and commenting on them to build a new text of their own.



After we collected and discussed these perspectives, I asked the students to develop their own definition of AI, using whatever pieces they had gathered that were useful, in groups at their tables. This same group of 8th graders (Mia, Susan, Jonah, and Trevor) produced the below chart paper, complete with robot illustration. They wrote, “AI (standing for artificial intelligence) can come in many different unique forms + variety of code & knowledge. Everyday examples of AI are google, facebook/instagram, Siri, tiktok, etc.” This definition does a lot of work, highlighting the broad scope of what might be considered AI (including “code” and “knowledge”). The group equates companies (e.g., Google) and applications/software (e.g., Siri) with AI. Even at this very early stage of instruction, students demonstrated how they were already incorporating a variety of perspectives into their understandings of AI. However, looking back on them as an analyst, these chart paper definitions still remained fairly abstract, and I think revisiting them nearer the end of the unit would have been instructive and potentially powerful.



As the unit progressed, students expanded their understanding of AI as present in other ways besides robots, and also deepened their understanding of how AI operates. Although students began broadening their views of AI early in the unit, this trend continued throughout. In week 5, Mia (8th grade) added “computers” to the common refrain of AI as more than robots:



This expansion is interesting as it implies a further broadening of the concept of AI as something that belongs solely to the disciplines of engineering/robotics and computer science, which was a goal of the curriculum. I wonder, though, what Mia meant by “much more”? Was she pointing to science fiction AI? Or the human aspects of it? Or perhaps she just meant it is in other technical objects, like phones.

In final interviews and writing from the end of the unit, students elaborated on their expanded and/or more detailed views of AI. Jimmy (8th grade) wrote, “I understand AI more, now I know a little about how it works. At first I thought it was just robots.” Jonah (8th grade) came to the understanding that AI “isn’t just creepy human like robots threatening to take over the world” and “I learned that artificial intelligence isn’t just robots but things like Google, Facebook, etc.” When I asked Theo (8th) “Have your views on AI changed at all?” Theo answered,

> To me, it used to seem so complex and now it seems like still complex, but in a more, like, complex in this way and this way, instead of just a general idea…and I think it's interesting to see like it doesn't feel as … like I would never be able to understand it.

All three of these responses demonstrate the broadened view of AI, but each highlights it in a subtly different way: Jimmy talks about knowing how it works, Jonah talks about an altered view of risk and also links AI to specific corporate entities, and Theo describes a changed personal relationship to AI as something now possible for them to understand. They still see it as complex, but in particular, rather than overwhelmingly general, ways. Theo’s response in particular pushed my thinking about AI as a complex phenomenon; I did not want students to only think of AI’s complexity, but also insert themselves into it.

However, some students also overgeneralized about what AI includes. This persisted from the first week—as when students equated Google and Facebook with AI in their chart paper definition—through to the final interviews. Susan wrote, when asked at the end of the unit how she defines AI now, “If it is an electronic…it likely uses AI.” Mona also wrote that “basically almost all electronics use AI.” Grey (8th grader) wrote, “I learned ai isn’t just robots, it is

everywhere.” June echoed this when answering where she encounters AI, with emphasis:

This is also an example of how the ideas of “what” and “where” AI is bleed together. Locating AI is one way we shaped our definitions of it, as something not just located in robots or science fiction, but as something that is also around us in our daily lives.

This was a pervasive trend, and students reported seeing AI “everywhere,” equated AI to phones and devices, said they were “tracking everything [they] do.” Adam defined AI as “something like Alexa or your phone.” While smartphones and many electronics do use AI, they are not equivalent to AI according to the disciplinary understandings of AI we explored. Juliet (7th grade) said AI was “everything,” but also followed that up with some examples in her final interview. When asked if anything in the unit will change anything about her life, she answered,

> ...I think that I'll definitely be looking out for like AI in the future…when I came in here, I like, when we started, I was like, "oh, AI, robots." But now I like know that it's all like, everything. Like when I go on my phone like AI's being used in like everything and it's kinda tracking everything I do. And I think that I'll like continue thinking about that when I move.

While I wanted students to see those less visible, and sometimes less physically apparent, ways that AI operates in our daily lives—in software, online, in “the cloud”—and they definitely did, I did not anticipate that they would start to feel like it was literally everywhere, pervasive and inescapable.

However, it is also difficult to determine at what point overgeneralization is “wrong,” and/or if this wrongness could be adaptive or instructive. Is Juliet’s assertion about AI tracking her an overgeneralization? Sometimes it can be hard to know. In class, we discussed the

reporting on whether your phone is actually listening to things that you say and the argument that, by combining search and location data from your phone and phones around you, advertisers do not actually need to listen to what you say to be creepily clairvoyant about what you have been thinking or talking about buying (Hunter, 2021). However, in some cases, your phone might be listening to what you say and using that information to target advertisements to you, among other things (Matarese, 2022). This fuzzy area, where it is unclear what the truth is and, beyond that, where you might find the most reliable answer, makes conversations about where the data that powers AI comes from very difficult. As we will see in the following section, it also complicates agentic decision-making around data practices and AI use.

There were also some things I did not know enough about to be able to say whether they use AI or more traditional programming (if AI is defined as a computer program that does not only follow pre-specified commands). This is unsurprising, given my own non-computer-science background and that technical functionality is sometimes hidden, purposefully (e.g., the "ghost work" (Gray & Suri, 2019) of people who are necessary for AI to function, but whose labor is concealed) and/or practically (i.e., via black-boxing). In these cases, I tried to be clear about the limits of my own understanding. For example, when the students were listing examples of AI and I was recording them on chart paper, I started a separate list for things I/we were unsure about whether they "used AI."

In summary, in the beginning of the unit, many students equated AI with robots. Although most students felt they had a better understanding of what AI is at the end of the unit, some overgeneralized what and where AI is—saying that it was everywhere, in (almost) everything, or equating entire companies with AI. This was a particularly difficult part of instruction, for several reasons: first, I *wanted* to complicate simple, straightforward conceptions of AI and demonstrate that it had become a part of our everyday lives; second, finding reliable,

current information about AI's presence and capabilities is difficult, so I often did not have enough information to untangle where AI is operative from where it is not.

Theme 2. Making AI "Two Ways": AI as Technical and Speculative Productions

In this section, I will discuss how I asked students to make "technical AI" using computer models and "fictional AI" using speculative thinking, writing, and drawing. I have divided this section into three parts, presenting data on (1) AI as technical production, (2) AI as speculative production, and (3) how these activities (failed to) meaningfully inform one another. Although I focus here on these two ways of approaching AI, and teaching about/with it, these are not the only ways we viewed AI; for example, we also discussed how AI is constructed through discourse, like in news stories and marketing materials. However, I focus on these two particular ways of understanding and creating AI because they are frequently set in opposition to one another (e.g., is this "real" AI, or just something from science fiction?) and because the computational-model version of AI dominates curricula, but the speculative production of AI also holds a lot of promise for meaningful engagement with AI, and AI and ethics in particular.

From the outset, I asked students to consider what might count as AI from different perspectives. I wanted students to see AI as both a technical phenomenon and a social one (i.e., a phenomenon created through human discourse and imagination). The first way I did this was by asking them to go on the Jamboard "Scavenger Hunt" mentioned in the previous section. I also designed activities throughout the unit that approached AI from a technical standpoint (as a combination of data and algorithms, as a computational process) and from a speculative one (as the product of human imagination and language). Ultimately, I wanted these activities/approaches to serve not only as different entry points into discussing AI, and understanding what and where it is, but also to bridge *between* the technical and fictional perspectives as the unit progressed, in order to deepen and complexify ideas about AI. I had hoped, as I write below, that having learned some details about how computer scientists define

AI and some key things about how AI works, that students might incorporate (or re-imagine, build on, etc.) some of these details in the fictional AIs they created.

In the following sections, I focus specifically on two activities I designed to understand AI from different perspectives, as a technical production, and as a speculative one. Beyond searching for examples and generating definitions, as we did at the beginning of the unit, I wanted students to experience "making" AIs in these ways in order to help them understand and define it. To draw on this understanding-by-making idea, I asked students to "build an AI" two ways: by using Teachable Machine to train an AI model to recognize objects and by creating a science fiction AI.

AI as Technical Production

On the second day of instruction for the 7th and 8th graders, we trained and tested image classifier models with Google's Teachable Machine, a browser-based tool for quickly and easily creating AI models (https://teachablemachine.withgoogle.com/). I used an idea from a Teachable Machine tutorial (Webster, 2019) to demonstrate training a classifier AI model to recognize the ripeness of a banana. Then I gave the students a variety of fruits and stuffed animals to train and test their own image classifier models. The classroom was bustling with activity as students tinkered with their models to get them to work as they wanted, and we explored ways that the models "break" in real time, together.

Although this was the only time the students actually "made" an AI by training, testing, and generating code, throughout the unit I also incorporated many other opportunities for "technical" exposure to AI. These included using materials from other extant curricula (e.g., Payne's AI + Ethics Curriculum, MIT's DAILy Curriculum), including playing a game where they embodied a neural network, learning about the functioning of Generative Adversarial Networks, and other activities. As discussed in Chapter 2, being a producer/author/creator is a key critical digital literacy practice, and so I built in ways that students could be a producer of a technical AI.

I also wanted to do this in as “authentic” a way as possible. Making their own functional model gives students a different experience of AI as a technical process than gathering definitions, than playing a game, or learning abstract concepts that undergird the process.

At the beginning of the session where they worked with Teachable Machine, I asked the 8th grade students what they thought someone would need to know to make an AI. Nearly every day of class, I began with a “pre-writing” prompt to get the ideas flowing while I finished setting up my recording equipment and teaching materials. I told the students to write as much as they could, and not to worry about it being good, just to keep their pens and pencils moving so that the ideas might flow. I asked this question in particular because I wanted to know, before they “made” their AIs, what they thought would be important. Would most of them emphasize technical skills? What disciplinary knowledge might they mention? Or perhaps they would take a collaborative view and think about distributing knowledge across a team? And what could the knowledge they describe as necessary reveal about their definitions of/implied models of what AI is?

Indeed, most students wrote that computer programming and/or knowledge of hardware would be important. Jonah (8th grade) wrote, “To make an AI, you would need to know the basics of computers, [programming] etc.” Mona wrote, “AI is a certain form of technology, because of this you’ll need to know some science & your way around a computer/phone/just technology in all.” Simon wrote,

> you need to be able to code for the AI to do anything or at least for robots with AI. To know how an AI works you would need a computer in the AI to send it signals to move. With an alexa it has [a] computer where you verbally send it commands.

Jonah, Mona, and Simon paint a picture of AI as a technical, computational thing that involves computer processes and hardware. Although these were the dominant kinds of responses, others centered human or animal intelligence as a model.

Two students thought having a good understanding of “animal” or “natural” intelligence was necessary to make AI. June wrote, “You need to be able to understand how Natural intelligence works so that you can [replicate] that knowledge into Artificial intelligence. You need to be able to replicate a lot of knowledge.” Theo wrote that you would need to know

> A lot. You need to know how animals think because AI thinks on its own without [“if this, do this”]. You can not come up with every possible outcom[e] of something that supposed to do something on its own. AI are [similar] to programming robots but robots follow the same steps every time. AI needs to be able to problem solve.

Theo here differentiates between “if this, do this” procedural programs and AI, highlighting how AI must “problem solve” because it would be impossible to program every scenario procedurally. This is a nuanced understanding of AI as something different from, and more expansive than, step-by-step programming, and perhaps closer to something like animal thought than to “robots.”

Students offered different ideas about where the knowledge or know-how to make an AI might come from. Mia thought that this “knowledge” would come from “education”, writing,

> I do feel like you would need to consider how much patience + time you have. Another thing to mention would probably have some type of coding/computer science knowledge. Without either of those types of education it will either add lots of time to your plate, or you won’t end up finishing.

On the other hand, Grey mused about whether you might just need to do “a quick Google search”: “I’m not really sure because I’ve never attempted making AI or trying to understand it. A quick google search may be enough, maybe not. I think it also depends on what AI you are trying to make or understand.” Although on first read, this may seem glib when considering that Grey was referring to one of the most advanced technological capabilities we have, it is in some ways very true; depending on the level of customizability and transparency you need or want in

your AI, a “quick Google search” really “may be enough.” We saw this in action when we trained Teachable Machine models in our browsers; Google “Teachable Machine” and you can find out all you need to create a working banana ripeness detector.

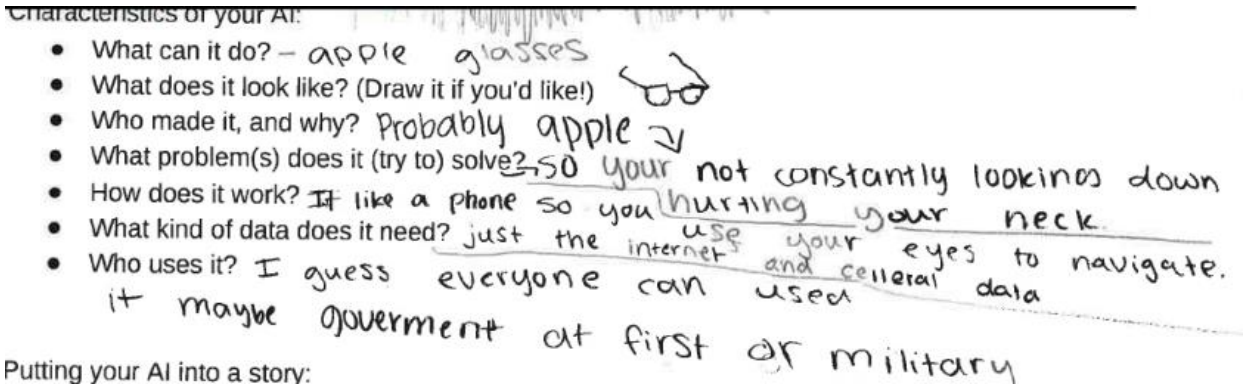
In summary, students saw knowledge of computer hardware and processes as important to making an AI, but had different guesses about how involved the process might be, from as easy as Googling or difficult enough to require a lot of patience, time, and education. In the next iteration of this design, I will ask students to reflect on this question again after they have experienced Teachable Machine, as I wonder if or how their perceptions shifted as a result of the exercise. Although I can see how the computational model version of AI traveled through the unit with them, appearing, for example, in their end-of-unit definitions of AI, I suspect their ideas shifted as they engaged in training, troubleshooting, and critiquing their own models. Since we did not look at the underlying code that Teachable Machine used, this also could have had the effect of over-simplifying the technical components and processes of AI, raising questions about what exposure to these things are “just enough” to be able to not only define, but critique and participate in the world with AIs.

AI as Speculative Production

Moving to the speculative mode of creating AI, I also asked the students to create a fictional AI and put it into a story (among other speculative activities). I wanted students to have another access point to creating AI that did not rely on computational expertise or materials, a way to explore the possibilities of AI that drew on the strengths they already brought with them—their capacity for creativity. This is not a new idea; Payne’s (2019) AI and Ethics for Middle School Curriculum also asks the students to create a future AI. However, I wanted the students to put their AIs into their own longer narratives, too.

The example below shows the questions on the “Build Your Own AI: Sci Fi Edition” handout I created to guide the activity for the 8th graders in week three of the unit. I had hoped

that these prompts would elicit some fictional inventions that would incorporate some of their learning about how AIs function, but on the whole the students gave only brief answers to my prompts. Using the example below, Hailey did not get into much technical detail when she described what kind of data her AI invention needs as “just the internet and [cellular] data.” Most completed stories did involve, however, failures of their AIs to function—from small glitches to going rogue—but students did not dwell on the details of the malfunctions, instead emphasizing the effects. This design failed to elicit deeper technical details about the invention that could connect the way an AI works (e.g., training data) to an ethical dilemma in the story. However, perhaps this failure is requisite for freedom to invent, and to gloss over details to make a good story. The strengths of these stories were not in the technology design, connecting fiction to the technical details they learned in the unit, but in the ethical experimentation the activity allowed. I will explore this function of the activity in depth in Chapter 6.

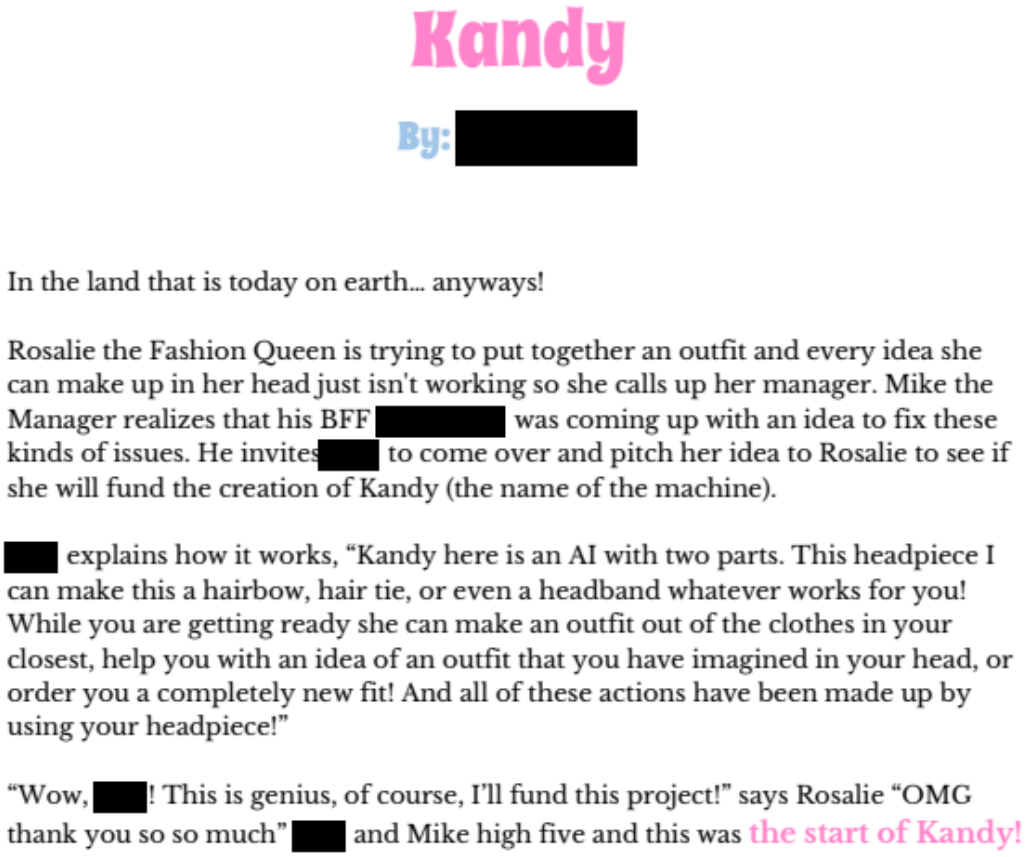

Characteristics of your AI:

- What can it do? – apple glasses
- What does it look like? (Draw it if you'd like!)
- Who made it, and why? Probably apple
- What problem(s) does it (try to) solve? so your not constantly lookinos down
- How does it work? It like a phone so you hurting your neck.
- What kind of data does it need? just the internet use your eyes to navigate. and celleral data
- Who uses it? I guess everyone can used it maybe government at first or military

Putting your AI into a story:

In the following story, “Kandy,” (student names redacted) Tori includes an entire paragraph describing how her mind-reading, fashion-generating AI works. Despite this, her AI origin story focuses more on the success of finding an investor to fund Kandy rather than exploring what Kandy would have to do to be able to access her users’ thoughts. Tori has interpreted “how it works” to mean “what it does” rather than a description of Kandy’s technical functionality. Although the activity did not include the technical richness I had hoped for, it did

other things: Tori writes herself into the story as the inventor of Kandy, giving her power both in the world of the story and outside it to develop her fictional AI. In an earlier activity, Tori had picked “stylish,” as an important piece of her identity, and Kandy clearly reflects that interest. While this activity did not yield the blend of the fictional and technical I had imagined, it allowed students to express their interests, author their own stories, and envision themselves as AI creators.



Kandy

By: [redacted]

In the land that is today on earth... anyways!

Rosalie the Fashion Queen is trying to put together an outfit and every idea she can make up in her head just isn't working so she calls up her manager. Mike the Manager realizes that his BFF [redacted] was coming up with an idea to fix these kinds of issues. He invites[redacted] to come over and pitch her idea to Rosalie to see if she will fund the creation of Kandy (the name of the machine).

[redacted] explains how it works, “Kandy here is an AI with two parts. This headpiece I can make this a hairbow, hair tie, or even a headband whatever works for you! While you are getting ready she can make an outfit out of the clothes in your closest, help you with an idea of an outfit that you have imagined in your head, or order you a completely new fit! And all of these actions have been made up by using your headpiece!”

“Wow, [redacted]! This is genius, of course, I’ll fund this project!” says Rosalie “OMG thank you so so much” [redacted] and Mike high five and this was the start of Kandy!

Putting the Technical and Speculative Together

Students told me they enjoyed both technically oriented and speculative activities, and writing fictional AIs and doing more computer-science-oriented activities both made it into their “favorite activity” answers at the end of the unit. Participants also described to me how they were making sense of AI as a computational object or process and as a fictional or imaginative one. When I asked Juliet (7th) if any of her opinions changed throughout the unit, she replied,

> when I thought of artificial intelligence I was like, oh, like robots, sci-fi, like Star Wars and stuff like that. But now I think of it more as like, oh, it's on a, like a, like datasets, like training it, like programming, and stuff like that. So I think they have changed, kind of.

I responded, didactically, encouraging her to think about it as *both* of those types of things. While I view my work as bringing my expertise—literacy and literature—to join up with computational understandings of AI, Juliet reminded me that it works the other way around, too: some (perhaps many) students already knew AI as a fictional object, but had little exposure to how it operates as a computational one in the current day. Furthermore, the activities we did seemed to have the opposite effect of bridging perspectives: Juliet saw her ideas of AI as shifting from fictional to computational, rather than including both. Instead of bridging, it seems that the effect of the unit was to supplant one perspective with another.

One final interview conversation stood out to me for its relevance to the tension between technical and non-technical approaches to AI, and the problem this poses for critical posthumanist literacy. Theo (8th) and I also discussed the relationship between computational processes and AI in their final interview. Theo told me,

> Yeah. Well um I guess to me, it used to feel like this vast thing of like I kind of think about the like ones and zeros and ones, like that type of stuff. But now to me, it's more like, "oh," like it's not just like magic-seeming, like there's stuff behind it. And I g—that's—I—feels good for me to know because it's like someday, you know I might actually be able to do something related to this, um and it feel like possible instead of like, when I was younger before I like just any coding, I was like, "Creating a website, impossible. I would never like be able to do something like that." But then like with simple tools, I'm like, "oh yeah, this is cool." And then I'm like I saw like what actual code looks like, that wasn't like Scratch. And I was like, I could never do that. But like I feel like AI has little to do with coding and stuff. So. Is that right?

I reply, discussing a no-code AI movement I had read about in a recent *New York Times* article, but also referencing the code they generated when they used the Teachable Machine models. I was delighted that Theo felt that they might be able to engage with AI, where before this had seemed impossible. I also see now how the approach I took sowed some confusion about the role of the "stuff behind it." Theo's experience—ping-ponging between seeing computational processes as impossible (code as 0s and 1s), then approachable (with simple tools like Scratch), then impossible again (seeing "actual code")—highlights the difficulty of moving between deep technical exposure and ease of access, a tension that computer science educators know well. The question that this raises for critical posthumanist literacy, though, is, what technical simplifications allow "just enough" access to the functioning—and potential for malfunctioning and/or misuse—without compromising students' abilities to effectively critique or create with AI? This remains a vital question, surfaced at the end of the unit by Theo.

Students had fun and valued (according to course feedback) the more technical—creating stuffed animal detectors/bananameters with Teachable Machine—and speculative—creating science fiction AIs—ways of making AIs. Although I intended for these to show how different entry points for defining AI could be valuable and mutually instructive—with the technical informing the fictional and vice versa, these activities seemed to reinforce the boundary between "real" and fictional AI rather than bridge it. One potential benefit of this is that students also were developing their abilities to distinguish between the current capabilities of AI and the feared or hoped for versions of AIs. However, in future designs, building opportunities for technical and speculative perspectives to build on one another could allow for students to strengthen both perspectives. Furthermore, these "two ways" are not the only two approaches to making, defining, or understanding AI. Bringing in, for example, more explicit discussion of sociological approaches to thinking about AI could disrupt the tendency toward strict binaristic thinking.

Theme 3. Missing Perspectives

Several students expressed an interest in exploring the history, development, or "invention" of AI. Theo (8th grade) wrote about wanting to know "How AI was discovered or the first AI." They had expressed this idea from the outset, answering the pre-survey question "What questions do you have about AI? What questions or topics relating to AI interest you?" with "who came up with it." Karl (7th grade) wanted to learn about "how they created and invented new and old technology." Sawyer (8th grade) initially made the focus of his final project a historical AI chess board, although switched his topic to AlphaGo later on. He was deeply interested in the origins of AI (and also one of our most vocal skeptics of the benefits of current-day AI).

I saw this echoed in one of the interviews with a teacher who was present for much of the unit. Although I focus heavily on student responses to my teaching, I sought the expertise, advice, and additional adult perspective of the students' regular teachers. The "adult perspective" turned out to be important in gaining insight about the role of a developmental or historical perspective, especially in the absence of formal computer science training. In reflecting on the unit, Mrs. Khoury shared how having seen the development of computing and AI happen gave her important context despite not having particular computer science expertise:

> Well, I think, you know, for people that grew up prior to the internet, that uh anytime I'm in a room and hearing something like this, I have all the background information. My first computer class was in maybe second or third grade at a small private school that had a computer lab, it was like four Apple computers that were giant and we got to write in soft—,uh, programming, and we made Asterix snowmen. So like, technically, as it relates to what you're asking… I understand, even though I don't know what I don't know about computer science, I do know a little bit about the evolution, technically and in our lives. Right. You know, what I understand is, you know that it's gone from the giant room full of machine machine parts with the punch-in cards sometimes to something in our

> pockets… Like, I would say it might benefit them, you know, do they really understand what non machine learning is? You know as a, as a elemental place to begin talking about AI, and then the, the technical aspects like, you know, understanding how that works technically versus how it wasn't a possibility before with what we were doing.

In this excerpt, Mrs. Khoury talks about how her personal experience provides a valuable resource for her understanding of AI as a development out of historical computer science. She sees it as something to draw from in the absence of computer science knowledge ("I don't know what I don't know about computer science, but I do know a little about the evolution, technically and in our lives"), and offers that this might be a valuable way for students to see AI as well. Clearly, the historical, developmental perspective of AI gave Mrs. Khoury a sense of understanding of AI outside of disciplinary knowledge in computer science; she knows it because she saw it appear, and change in her own lifetime. Perhaps we can learn from Mrs. Khoury's experience here and take up her idea of sharing that information (in the absence of being able to share experiencing this change firsthand) as another entry point for students to understand AI and its role in our lives.

Beyond missing a historical perspective on AI's development, one student also looked forward, wondering about how "what AI is" will change as more people become involved in shaping it. In response to my final interview question, "Are there things about AI that you're really excited about or you think that will like be big benefits?," Susan (8th) responded,

> …well I think everyone should learn more about AI so we could all have an understanding how it works. But I feel like, it's just kind of like what would happen if everyone like kind of learned more about that to see if that would make a difference on what it is?

Susan is wondering here how bringing in missing perspectives might change the meaning of AI itself.

In summary, one key perspective that was missing—and missed by students—from the unit was a historical one. Students were curious about how AI was made and who made it, and one of the teachers described how having experienced the development of AI in her lifetime gave her important context, in the absence of computer science expertise, and that students might benefit from understanding AI's evolution as well. This is both a failing and a success of the unit; a failure because I omitted an important perspective that could have provided rich context for understanding what and where AI is, and a success in that the students were exercising important critical literacy skills by asking whose voices and viewpoints were left out. This latter point became even clearer when Susan mused about how AI will change once those missing perspectives become a part of the trajectory of future AI.

Thinking with Data and Theory for Redesign

In my design, I drew on my own disciplinary experience in literacy—first as a librarian and now as a literacy researcher—and on my exposure to posthumanist theories that emphasize complexity, dynamism, and interconnection of things. For example, I called AI a "phenomenon" in my materials and in instruction, and I emphasized its many instantiations in fiction and in the current-day world. One way that I failed my commitments to posthumanist theory, however, was largely ignoring the material aspects of AI; though I discussed the role of server farms and the movement toward sustainable AI toward the end of the unit, I clearly privileged computational and human processes over anything like Bennett's vibrant matter (2010). This, like the history of AI, is an important missing perspective that could influence how students conceptualize AI. Even the way I asked the question about what you need to know to make an AI obscures material requirements, as I privileged "knowledge" here at the expense of material considerations; I could—and perhaps should—have also asked what materials or agents they would need to build an AI.

Interestingly, although I set out to work across boundaries between "real" (currently operable) and fictional AI, my teaching served to reinforce these boundaries. Here I can think with Barad's (2007) idea about how we produce phenomena through a process of "cuts." It makes sense that students would be interested in distinguishing between what an AI can do now and what it cannot, to form the "cut" there between possible and potential, so that they can evaluate the risks and benefits to them. The thing that I would like to design toward, though, is exploring how we are constantly producing, and re-producing the meanings of AI—real and imagined—from within it; we determine what "counts" as AI alongside the sociomaterial systems that enable and constrain us (the subject of the next chapter).

Chapter Conclusion

How we define and locate AI is consequential for how we live with—and teach about—it. By blending definitions from different perspectives and encouraging students to notice their everyday uses of AI, I hoped to provide broader access to the concept beyond what students might get in a computer science class. In this chapter, I discussed three themes I saw when analyzing data relating to students' (and my own) understandings of what and where AI is. These trends are impetus for redesigning toward more opportunities for clearly delineating—then disrupting—the boundaries between technical and imaginative instantiations of AI and for incorporating missing perspectives (and encouraging students to continue advocating for those perspectives they would like to incorporate).

Chapter 5: Big Questions about AI: What Can and Should We Do about It?

This chapter centers issues of agency, ethics, and justice. Because these are deeply interconnected, I have included them as two parts of the same chapter. However, I present my findings on these topics in sequence, first examining research question three: (How) can this design and my teaching deepen students' understanding of their agency as entangled with AI systems and open up the potential for agentive participation in/resistance of these systems? and then moving to research question four: How can I design curriculum and teach to foster ethical engagement around everyday AIs?

Part 1: Entangled Agency and the AI/Data Audit

This section addresses my second research question:

> (How) can this design and my teaching deepen students' understanding of their agency as entangled with AI systems and open up the potential for agentive participation in/resistance of these systems?

This broad question covers some of the specific things I wondered about students' relationships with AI and how they felt about them, including:

- How do students connect themselves to AI systems?
- How can data audits support noticing students' connections to everyday AI?
- How do students express feeling about these connections and their abilities to choose or change them?
- What do students think about how they are surveilled and how their data is used?
- Reflecting on instruction and thinking with this data and theory, how might I redesign around the second instructional goal (students will explore personal connections to AI systems, including possibilities for agency and action)?

This section focuses on students' completed audits (see Appendix D) and on students' expressions—in discussions, activities, and interviews during the unit—of their own power to effect change in their uses of/by technology.

I have selected four themes to discuss in relation to these aims and my research questions:

1. Students saw AI as something they use, not necessarily something that uses them, and sometimes downplayed the value of their data.
2. Students expressed few concerns and also few opportunities for action around their data uses in their audits.
3. Privacy and safety were competing interests in student rules around data use, but students tended to agree that consent was important and selling data was disagreeable.
4. Students often made negative evaluations of their own power as young people, but demonstrated agency in other ways.

I then return to thinking with my theory and the data, exploring more pointedly posthumanist applications of agency in conversation with the data I discuss here.

Selected Themes in Locating Entangled Agencies

Theme 1. Students Say They "Use AI" More Than It Uses Them or Their Data

At the end of the unit, I asked the students to respond to the question "How are YOU connected to AI?" Nearly everyone wrote some variation of "I use it." Some added daily/every day or specific tools (Xbox, phone) or apps (social media, YouTube) to their response, but the responses were overwhelmingly consistent. Not a single student mentioned how AI uses them or their data. This is not to say that students did not see AI as agentive—this came out in other ways—but that the dominant way they felt connected to AI was as a human in control of a technological instrument.

The absence of responses about AI using or guiding the students was surprising given how much time we spent talking about the ties between data (including surveillance data collected of them, and data they generate) and AI. “Data” was a key concept for students in the unit from the outset, alongside “AI” and “ethics.” In the first session, I used a slide deck from MIT (Payne, 2019; MIT Media Lab Personal Robots Group & MIT STEP Lab, 2021) that included it as one of three key components of AI. I continued to emphasize the connection between these two behemoth concepts as we discussed how we all generate data as part of our everyday lives. In addition to this technical perspective—where I explained that AI needs the data we generate to function (train, predict, and improve), and that the who, what, and where of this data is impactful for ethical reasons—I also wanted to connect to critical practices and privacy. Data privacy is a topic close to my (librarian’s) heart, and I designed this AI ethics unit to incorporate learning about data and its circulation as a key part of understanding where and how AI is operative in students’ everyday lives. However, as I saw in the implementation of this unit and discuss in more detail below, tying AI, data, and agency seamlessly together proved difficult.

Students both underestimated and overestimated their data’s reach and utility. Jonah (8th) searched some things while playing a game, and in the critical evaluation column wrote, “There isn’t much they can do with a 14 year olds data.” Sawyer (8th) filled out a row in the audit about using TikTok, saying that he approves of TikTok’s use of his data “because they are not [getting] my personal data.” On the other hand, June wrote an audit entry for texting her friends, noting that “Apple records all my texts…maybe fo[r] security,” and that she does “not complet[e]ly” approve of this because “It kind of invalidates my privacy.” However, evaluating these responses—and then redesigning my instruction around them—ranges from difficult to impossible. This is because it is difficult to determine what is “true” in terms of corporate data privacy practices. We rely on partial (e.g., from journalists who investigate corporate practices), often unintelligible (e.g., in terms and conditions documents), and constantly changing

information when making these determinations. Although coming to the correct answer is not always the goal, and I—out of necessity and theoretical/pedagogical commitment—made questioning and researching an important part of this process, it still remains important. While it may matter more that students develop the critical skills that apply to a broad range of privacy questions, understanding how our data is generated, used, and sold is also consequential. This is a case of both content and process/skill being important. Given these confusions, it makes sense that students did not talk about how "AI uses them" or make clear connections between the way their data circulates and the operation of AI at a grand scale.

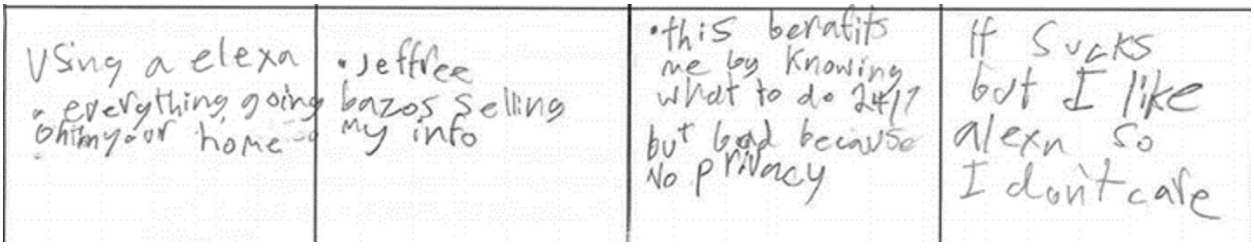
Theme 2. Students Express Few Concerns, and Also Few Options for Taking Action in the AI & Data Audits

In this section, I focus on the student AI & Data Audits. I chose the audit because I designed it specifically to encourage students to notice and critique their uses of and by AI during their daily rounds. The audit included a brief set of instructions (see Appendix D for a copy of the audit handout) and a table to fill out, with columns for identifying where and when they saw AI or generated data; who was collecting the data or using the AI and why; making a critical evaluation of the benefits and risks involved to them; and brainstorming possible actions should they disapprove of this use of/by AI. In these audits, students identified various entertainment, school, transportation, communication, and banking functions where data was collected on them in the course of their day.

Overwhelmingly, students saw benefits to how their data was being used to provide them with services that were valuable to them, and most had few, if any, concerns about how their data was collected and used. Most approved of data uses they identified in their audits, emphasizing the benefits to them—from saving money to seeing where their parent was via a location-tracking app—even if they could not control it. Juliet (7th) wrote, "I do [approve] of them [YouTube] using my data and this is out of my control because I can't stop them from doing it."

She concluded in her reflection, “I did see what I expected to see, I think I’m becoming more aware of where my data is being used… I think a lot of AI is used to benefit me and I don’t really have any concerns about security/privacy.” Students praised the apps and services that they used: “Yes I do [approve of this use of data] because i love spotify” (Adam). Bea wrote that she “would opt out if it [google search] sucked,” showing the cost-benefit analysis that so many of the students made; they knew their data was being harvested, but decided to continue using tools that they needed and/or liked.

A small minority of students were more grudging with their approval. Liam (7th) put it bluntly (pictured below): “it [“Jeffree bazos selling my info”] sucks but I like alexa so I don’t care.” In their reflection, Liam summarized, “I like AI and beli[e]ve that is the future [therefore] I’m happy it’s evolving. Here Liam’s responses show how they are simultaneously critical of specific data practices they see as an invasion of privacy, but also unwaveringly hopeful about the promise of AI on the whole. It is also interesting that Liam pins the blame on a particular person, Jeffrey Bezos, as the culprit who is selling their data. This is counter to the way I conceptualize agency (and, relatedly, blame) as emerging from human-nonhuman systems, distributed among people, social systems, and technological agents. I wonder if villainizing a single person has some important function for Liam, though: does it help justify their continued use of Alexa? Does it make it less scary, to think about a single individual at fault rather than an overwhelmingly large and entrenched set of systems and materials?



The embrace of the AI-enabled tools they encountered and the overall lack of concern over their data circulation was nearly universal on the audits. Only one student was consistently

concerned with the ways her data was circulating. She reflected, “It’s shocking how much data is being stolen. Its easy to know what AI is its [difficult] to find out how to opt out of this data use. I find it scary how much data is being given away. Privacy is being endangered by this, I think that apps should not sell your personal information because thats your personal information” (Hailey, 8th). This response—strongly affective, with “shocking” and “scared” and “endangered” hammering home her fear—more closely echoes those I saw during the previous studies in the larger design project, and I was surprised that Hailey was the only person who expressed these strong views in the audit.

Although these kinds of overwhelmingly negative and harshly critical responses were rare on this assignment, students still expressed shock in class discussions around how much data companies collect (and sell), especially on children. Students’ concern about their data privacy came out in other ways. For example, when reflecting on the unit at its conclusion, Mona (8th), said that “Be careful when sharing info” was the most important thing she learned. Bea (8th) also felt that her actions might change based on what she learned about AI being “in so many places,” and that she “might not do certain things….I might not click on something or I might be more like protective I guess about who my data is given to.”

Perhaps the difference in their expressions of (dis)approval is linked to the way that the audit asked the individual student to consider their own uses of and by technology; would I have seen different responses if I asked them to think beyond their own personal habits and choices? In the evaluations they made in the audit, they saw great personal benefit to using the AIs they identified, and very little personal risk to using them—even when they found the data practices of the company responsible distasteful in general, or harmful to others. I saw this, for example, in the case of Liam, who disagreed with the infringement on their privacy but still was willing to use Alexa and had faith in AI’s potential. In future iterations of the audit, I would like to see if

broadening the scope of their risk-benefit analysis, asking them to think about impact on others beyond themselves, might change the way they evaluate, or maybe even interact with, AI.

Little Desire or Options for Acting, Other Than Opting Out

I also wanted to know what students saw as their options for taking action in response to their evaluations. Some students wrote that they adjusted their privacy settings or chose how to share their information to control their data. Jonah (8th) wrote, “Im fine with it since I don’t post it on the [digital art] app.” Sawyer wrote, “No [I don’t approve of this use of my data], so that’s why I block it [snapchat location collection].” These types of responses, where I saw how students made informed, active decisions about what information they wanted to share, were rare on the audit. This does not necessarily mean that the other students do not make these kinds of decisions frequently, but they certainly did not include them on the audit. Most students offered opting out of using a tool entirely rather than using the tool in a different way than the default.

This raises questions about whether this reflects a flaw in the tool and/or an absence or desire to make changes to their data sharing practices. Because students overwhelmingly wrote that they approved of the ways the tools they identified used their data, focusing instead on the benefits to them, I am inclined toward the latter interpretation. I theorize that, because the tool encourages the students to think only about themselves, and their own individual uses of and by data, students’ responses draw on their individual threat models (e.g., how could this harm me?) rather than more abstract ideas about what is right for everyone, or others. I think the tool is effective for noticing everyday AI, and how we are often generating data that powers it, but less effective at putting this into the context of entangled—with other people, with nonhumans—agencies and ethics.

Theme 3. Making Their Own Rules about Data Use: Seeing Tensions between Privacy and Safety, Emphasizing Consent, and Disagreeing with Data Sale

In contrast to the audit, which was centered around individual connections and responses to AI systems, other activities prompted students to think about data practices more broadly. Before they began the audit, I asked both the 7th and 8th grade students to write their own rules for how they thought their data should be used. Although this activity creeps into the realm of the ethical and just, I include it here in the agency section of the chapter for a few reasons: first, it relates directly to data privacy, rather than overarching ethical issues surrounding AI, which is a focus of this section; second, I believe that asking students to develop their own rules is an agentic practice; third, by asking students to consider when and why their rules might not apply, they must also consider ways that agency might be conditional upon, as Simon puts it below, "different types of power."

In the 8th grade class, students developed their rules in table groups and shared them with the class. I then asked them to come up with an exception to their rule and share that as well. The table below shows their (transcribed) responses to this exercise. In the 7th grade class, I asked them to individually write "What rules should companies, parents, and you follow when using your data" as one of their pre-writing activities to start a day of instruction.

By asking the groups to reconsider their rules and generate an exception to them, I wanted the students to see that there are always ways that life will complicate our best-meaning attempts at making things theoretically just. With the exception of the fourth group, these students saw tensions between their initial sets of rules, which mostly erred on the side of protecting privacy, and ensuring safety or preventing illegal behavior. (The fourth group decided that surveillance in public places was acceptable.) By surfacing these tensions, the students practiced taking perspectives on what their systems of ethics should allow, playing the role of privacy and safety advocates as they tested out the limits of their rules.

This activity yielded different kinds of evaluations than the (subsequent) audit did. In these discussions, students had to take perspectives not rooted in their own experience or individual assessment of risks and benefits. However, both activities did yield similar practical implications for privacy vis-à-vis safety—although in the abstract, students wanted companies to use their data responsibly, and with consent, the end result was still overwhelmingly accepting attenuated privacy. In the student codes, this looks like allowing parents to “look at your history, your messages, if they have a suspicion that you’re doing something illegal” or allowing surveillance in public spaces. In the audit, this looks like continued use of tools with questionable data privacy practices.

Table 4

8th Graders’ Rules and Exceptions for Data Use

Group	Rules	Exceptions
1	Simon: We made one rule for each since they all have different types of power. For parents, they should be understanding of what you search and they should not be that nosy into searching your history. Companies, they need consent to use your data. Like, and they should be—you should be able to read the terms of service. Government, I think what they're doing is completely fine already. And that’s all.	Simon: This was under our parent section. They can look at your history, your messages, if they have a suspicion that you're doing something illegal.
2	Mona: Okay, so for companies we said like only use the data that they've gotten themselves rather than getting it from like an other type of um like source [Sarah: uh huh] and then only some recommendations like only— if you're like subscribed or have consented to it and then for the government like only use information that they have if they need to. and then for parents only really use it for safety concerns, but it would be nice to know if they were like looking in your search history.	Mona: Okay, so our thing like was, parents can only like use the thing for safety concerns. But our thing is something that the child is viewing that can be extremely dangerous…

3	Hailey: Um. If it's not your data you shouldn't share it because it's not yours in the first place. If you have someone's data, you have to ask to use it. Terms and conditions need to be shorter and be able to understand it. If you write it in [inaudible] only a lawyer could read it that's not like…[makes face] you should always have control of taking back information that you put out there and parental, uh parental, parental uh stuff. When it comes to like—They shouldn't have like full control of it but they should have control over an extent. Probably like the younger age, not like at this time, because you could get addicted to something that you shouldn't be addicted to [inaudible].	Tori: So our first one was that you shouldn't share other people's data. But if it's a situation, someone committing like a serious crime, I think you should probably Hailey: If you like know where they live [inaudible] Tori: Say that information to the police. Like if someone murdered someone you can go to the police and be like hey like I witnessed this crime, they told me about it. This is not okay.
4	Grey: Our four were: don't tap into our cameras, like, like our webcams, other stuff. Uhh. Don't steal info through private accounts, like a google account, icloud, stuff like that. Uh. Don't share or sell our location without consent. From, from the user, not the parent. Uh. And then, well…yeah. Bea: Make simple to understand, like, terms and conditions, private policies. So that we understand what we're getting into.	Theo: So. What we picked is um for like no, no recording stuff, in public areas that's okay. [inaudible]

This tradeoff between privacy and safety reveals important things about students' ideas about agency, when defined as the power and ability to act. Sharing data is an agentive practice here; revealing information, like in the case of group three, is deemed necessary when a life is at stake, or has been taken. I was surprised, too, by students' (in groups one and two) willingness to share power over data with their parents, or parents in general. A final theme that I will point out is the importance these groups put on transparency and consent; understanding and agreeing are also agentive practices here. For example, Hailey's group rules stipulated that

> If it's not your data you shouldn't share it because it's not yours in the first place. If you have someone's data, you have to ask to use it. Terms and conditions need to be

shorter and be able to understand it…you should always have control of taking back information that you put out there…

This group emphasizes consent, understandability of terms and conditions, and also continued control over the circulation of your data. In their formulation, consent is continual and conditional, not just a matter of getting approval in a single moment. They believe that an individual should be able to revoke consent as well as give it.

Several groups also address the differences in power among different groups, like parents and children, but no more explicitly than the first group, who decided to design their rules based on the understanding that parents, companies, and the government wield "different types of power."

While the 7th grade responses echoed this tension between ensuring privacy and safety, they also expressed some thought-provoking ideas about what companies should be allowed to do with data. I was struck by their acceptance of data being generated and collected, but not sold. At least four students explicitly included language around this boundary. For example, Cal wrote "Shouldn't sell information, just collect it." Patrick added an acceptable use of the collected data, saying that "Company can collect data for [recommendations] but should not [monitor] it." It seems that these students approved of data collection (by companies), but had firm beliefs about what was okay to do with that data. The data being gathered/existing or even being possessed by companies was seen as unproblematic, but taking the next step of selling it or monitoring it was where the problem arose. This could be a springboard for a deeper conversation about generating and consuming data, and in particular looking at opposing viewpoints about whether the (most pressing) problem is that these new, often intrusive data exist or that bad actors will misuse them. Connecting explicitly to agency, where does the (individual, human's) power to act become attenuated? Should we be able to intervene at the point of data generation, collection, and/or use?

Theme 4. "As a child, I don't think that I have that much power": Negating and Affirming the Power to Enact Change

I was deeply interested in how students perceived their own human agency, or their felt ability/capacity to take action or enact changes in their own lives and the lives of others. Being so thoroughly steeped in critical perspectives on AI, where bias and oppression are focal topics, I was also hoping (though I know it is not fair to place this burden on them) that the students would help me see the way to infuse more hope into this kind of instruction. I did not want students to leave my class or the unit feeling scared or powerless, even though sometimes these felt, to me, like appropriate responses to certain current and potential applications of AI.

I was surprised by how bluntly the students described their feelings of powerlessness and negative views of children's knowledge or abilities throughout the unit. During an 8th grade whole-class discussion on day two of week two, we debriefed after reading an excerpt of UNICEF's children's data governance manifesto (Byrne et al., 2021). Tori had just raised the point that a minor had the most data transmitted to more companies than any other people they looked at in a study. In response, June says, "I think that's because minors aren't as smart as adults, usually." Another student interjects, "they can't DO as much." June agrees, "yeah. so they can—they don't really know what buttons to press and what ones not to." I was startled by this attitude, though I did not challenge it in the moment. How had these students internalized such deficit views of their knowledge and capabilities?

Responses like this cropped up in several end-of-unit interviews as well. In their final interview, I asked Dom (7th), "how [do] you feel about the power that you have to make changes about the way technology works in the world?" Dom replied,

> Um. Well, I mean, as a child, I don't think that I have that much power currently, but I mean, in the future, if I would like to get into a job in computer science or something along the lines of where I would be dealing with AI and ethics, I think that that would

have quite the power to change things. But currently, um, I don't think I could do much unless, you know, I get on some really ...thing and I go viral. I don't know. I don't, uh, I don't know if I'm cool enough for that yet. I don't know.

Dom saw themself as potentially having power in the future, except if they happened to "go viral." This was a surprising view of their entanglement in these systems, and reminded me that virality might be an important tactic in modern social movements, and Dom certainly saw it this way. In Dom's view, being "cool" meant possibly going viral, which meant having power to make change. It is unclear, however, whether Dom sees "going viral" as a joint accomplishment of the software/algorithm/AI and humans, or whether it is only a matter of "cool," or social capital.

However, Dom ends up talking themself into feeling like they have "a little bit" more power:

> Sarah (6:26): I don't think it's about cool.
>
> Participant (6:36): Yeah. Obvi—I mean, no. But I mean, if somebody wanted to listen to a plan that I had, that would be cool. Like, but I don't think I'm going to be able to convince a bunch of governmental officials to do something or another about AI. I mean maybe, you never know.
>
> Sarah (6:57): Yeah.
>
> Participant (6:57): Actually, you know, I, there's like, yeah, there's more of a, now I think about it, there have been a bunch of kids to do that, so yeah. I actually think, yeah. I think I do have a little bit. Yeah.

Upon further thought, Dom remembers that "there have been a bunch of kids to do that," presumably referring to "convinc[ing] a bunch of governmental officials" to take action. Perhaps using examples of successful youth campaigns would be a powerful tool for instruction, so that other students could draw on these examples like Dom did.

Having little power was not only linked to youth, but also to the relative strength of collective action. While June described feeling a personal sense of agency, she qualified this by minimizing the impact that her choices alone would make. When I asked her, "do you think you

have options and choices, and that you might be able to shape the way that AI is developed as you start using it more and more?" June replied,

> Yes. But it won't have a big impact because I'm only one person. So if I stop using certain websites because their privacy policies are crappy, then it's, if I'm the only one doing that, it probably won't make a difference. But if a bunch of people do it, it'll make a big difference.

I include this response especially to point out June's recognition that her agency is not just entangled with machines, but with other people. This tension between individual and collective power and action has existed for millennia; in the age of AI, how does the further entanglement of machines affect this balance/tension? As we saw in Dom's response, it is not negligible: the concept of "going viral" is a case example of this shift. "Going viral" is a joint human and machine (collective) action.

Students also described constraints to their ability to choose a particular course of action because of school requirements. In response to the final interview question, "What do you think are the biggest ethical challenges about using AI?" Susan (8th) said,

> Um I think that, just like kind of sharing your information to other people, it could be bad in some situations. Like, as we're talking about, and like if you like put your stuff out there then the websites will take it and use it as their own ads, like that kind of thing, um and so it's just kind of like...But you kind of have to do it, so it's like if I like had like to do stuff for school and I had to put my info out there, but then it had that problem, I'm like well I can't do, I can't not do that.

Susan "couldn't not" share her private information if school required it. This highlights the tremendous responsibility that teachers and curriculum designers have for ensuring that technology use protects students' data privacy, as students do not have—and/or do not feel they have—the ability/power to opt out. For example, I tried to only use tools in my instruction

that were minimally invasive, data-wise, checking privacy policies and consulting Common Sense Media's (2023a) privacy evaluations (https://privacy.commonsense.org/) before selecting a digital tool.

On the other hand, students frequently shared their AI expertise with me and others and sometimes did assert their ability to make change. Dom in particular was excited to teach me about AI singers called "vocaloids" that I had not seen before. Some students described ways that their new knowledge of or experience with AIs throughout the unit might give them new ways to participate in the world. When I asked Dom, "So what do you think the most important thing you learned during the unit was?" they replied,

> What ethics was. Because I didn't know what it was and the whole concept of it before I came here and I think learning like about prejudice in AI, because I never had thought about that. I think those are two really, really important things that I learned about…And like now that I know about them, I can take responsibility in like clearing that up and stuff.

Dom makes a clear connection between ethics and agency, the two foci of this chapter, in this response. Their new knowledge of ethics and "prejudice in AI" together allow her to "take responsibility" for working toward social good.

When asked if she thought anything she had learned in the unit would change what she does in her everyday life, June answered,

> Mm-hmm. Because I know that… I, I think, I think it's cool, because I did a project on advertising AI… So I know a lot about advertising now. And I think I'll know how to get better ads, or ads that work for me more.

June connected her knowledge of how AI-powered advertising works to being able to shape the ads she sees; her awareness and knowledge could allow her to take steps to take advantage of this privacy-invasive practice.

Thinking with Theory and Data for Redesign

In this findings section, I discussed how students felt about their own power in the world, and their capacity to act should they decide it was necessary. I want to dwell here on linking this data directly to posthumanist conceptualizations of agency. Barad (2007) insists that agency is not something one possesses, but is performed dynamically. If we think through these lenses, perhaps it makes more sense that students would equivocate about their feelings of power. For example, Susan (8th) said, “I feel like every time I see like a terms and services or even recently I'm like ‘I learned about this,’ like I know that this is [inaudible] AI and stuff, and like all of the apps like and stuff they use AI.” However, when asked if this would link to a change in action, Susan said “I think it's, it's more of an awareness, kind of, but I definitely give some thought to it, and like it depends on when I'm in a certain situation, so... [inaudible] I kind of have to encounter it to know.” Here, Susan reminds me that her agency is produced variably depending on the context—very Baradian. She recognizes that her knowledge does not automatically translate into action, or the ability to act, but rather is context-dependent and emergent (produced when she “encounters” it). Susan does not “have” agency, but it is produced in the moment, and knowledge is necessary but not sufficient for taking action.

I also draw on posthumanist theory, and Morton’s (2013) hyperobject in particular, to make sense of the disconnect I saw between their heightened awareness and their continued resignation about their ability or desire to make changes. In their final interviews, I asked students if participation in the unit might change anything about their lives, or if anything surprised them. Their responses surfaced fraught relationships between awareness and action. Grey (8th) was surprised by “...how much there is AI everywhere and like data is being taken without you knowing. And even if you know, it's still being taken.” Here, Grey describes a certain futility of awareness. Morton’s (2013) explanation of the hyperobject gives me a lens to consider this situation. Morton describes how, when faced with the vastness of a hyperobject it “hollows

out my decisions from the inside. Now every time I so much as change a confounded lightbulb, I have to think about global warming" (124). Add to that Morton's idea that there is no right way to act from within a hyperobject, and this is a powerful way of explaining why students saw AI everywhere but did not feel the need or ability to do something different with this knowledge. Perhaps their decisions had already been "hollowed out" by the enormousness of AI.

Posthumanist theories of agency can help explain these responses, but what does it mean to design for/with them? This raises more questions than answers for me. In practice, does it have to mean showing students how their choices are shaped—most often constrained—by AI, as we did when we learned about TikTok's algorithm, or when June researched AI advertising? Does it necessitate acknowledging the hollowness and wrongness of any individual decision we make, like Morton's (2013) example of changing a lightbulb in the midst of global warming? If the effect of knowing, of awareness, is resignation—like we see in Grey's audit response of "even if you know, [your data] is still being taken"—(why) should we still teach students to be aware? Can we think of awareness, in itself, as agentive?

Section Conclusion

The connections among data, AI, and agency are complex, and the audit was just a beginning step in surfacing them. A key question that arises from this analysis is, Is awareness, by itself, agentive? While traditional conceptions of agency are action-oriented, perhaps posthumanist theory can help us rethink, without discounting the importance of action, the relationship between awareness and agency. For these students, it might be one of few options for experiencing a (albeit limited) sense of power in the face of a big, looming web.

As a final note, although I focus here on data audits, and data privacy, we interacted with AI in many ways throughout the unit: training an AI model to recognize fruits and stuffed animals with Google's Teachable Machine, interacting with an AI Holocaust survivor, and inventing

fictional AIs to put into stories. All of these entangle us with AIs, as we interact with, build, and generate data for them.

Part 2: Ethics & Justice: How Should We Live with AI? Using Ethical Lenses to Think about AI, Society, and the Future

In this section, I address the overarching research question:

How can I design curriculum and teach to foster ethical engagement around everyday AIs?

In particular, I investigate the use of formal ethical theory and discussion of meta-ethics in my instruction, asking,

- How did these students, in response to my curriculum and teaching, learn to use formal ethical lenses, especially to generate ethical questions and opinions about AI?
- What ethical issues relating to AI held the most import for students?
- (How) did discussing meta-ethics deepen understanding of ethical ambiguity and complexity?
- Reflecting on instruction and thinking with this data and theory, how might I redesign around the third instructional goal (students will practice thinking ethically with respect to uses of and by AI systems and develop their own ideas about what is ethical with respect to AI)?

As I was designing the curriculum, I debated about how to organize the discussion of ethics. I saw two major approaches to organizing instruction around AI and ethics: through ethics issues/topics and through ethical practice/thinking. My experience in the previous phases of the project gave me different ways to approach the subject. The undergraduate coursework I observed was organized around AI ethics topics (e.g., AI in healthcare one week, AI in education another) rather than around ethics as a more general set of theories and practices. I wanted to try out a practice-based rather than topic-oriented approach, given the relatively short

amount of instructional time I had with the students. I was also hoping that this might help foster an ethical thinking practice, where students would get in the habit of seeing right and wrong from different viewpoints. I hoped this would be more flexible than developing topical expertise in the short time I had with the students. I also hoped that seeing ethics as a process of inquiry and perspective-taking would give students permission—and some tools/language—to question why we might think of certain things as wrong or right in the first place (venturing into meta-ethics).

However, I still tried to balance the introduction and application of normative ethical lenses with discussion of ethical issues specific to AI. To get at both of these goals, I incorporated work around normative ethical lenses, including direct instruction about major normative theories, opportunities to practice applying different lenses to both hypothetical and current-day ethical dilemmas, and around major topics in AI ethics, like focusing a week on AI & environmental impact/sustainable AI. I also wanted to see how speculative fiction writing could provide an avenue for ethical experimentation, but I will discuss that separately in the next chapter.

Selected Themes in Exploring Ethics

Theme 1: Students Used Ethical Concepts without Being Dogmatic about Terminology.

I included direct instruction about dominant ethical theories and meta-ethics (for the 8th grade only due to time constraints), giving short slide presentations on these topics before asking students to apply them in various activities. At the end of the unit, I asked students directly whether they thought that learning about various dominant ethical lenses, including deontology, utilitarianism, and virtue ethics, helped them. Some of them also volunteered "ethics" or "different kinds of ethics" when I asked what the most interesting or important thing they learned was, at the beginning of the interview. Bea was one of the latter, sharing that "I

didn’t know there were types of ethics, I guess…I thought it was just everyone for themselves.” Simply knowing *that* there were different perspectives, not even getting into what each one was, proved helpful to Bea.

When asked about whether knowing the names of theories was helpful, June said, “I think it's good to have those terms to sort, like sort all your thoughts.” Grey offered that knowing the different perspectives can both help you clarify how you think and also help you explore other perspectives:

> If you don't know how you think, but then you see one that's there, and it's like "Oh! That's how I think" then you can [inaudible] Then you can say, "Oh I think this way." Or you can try out different ones so that you experiment to see how other people would think, or how you could possibly think.

Grey saw both the reflective and prospective value of learning the different lenses, as she could classify the way she thought and also try out new ways of thinking as a perspective-taking “experiment.”

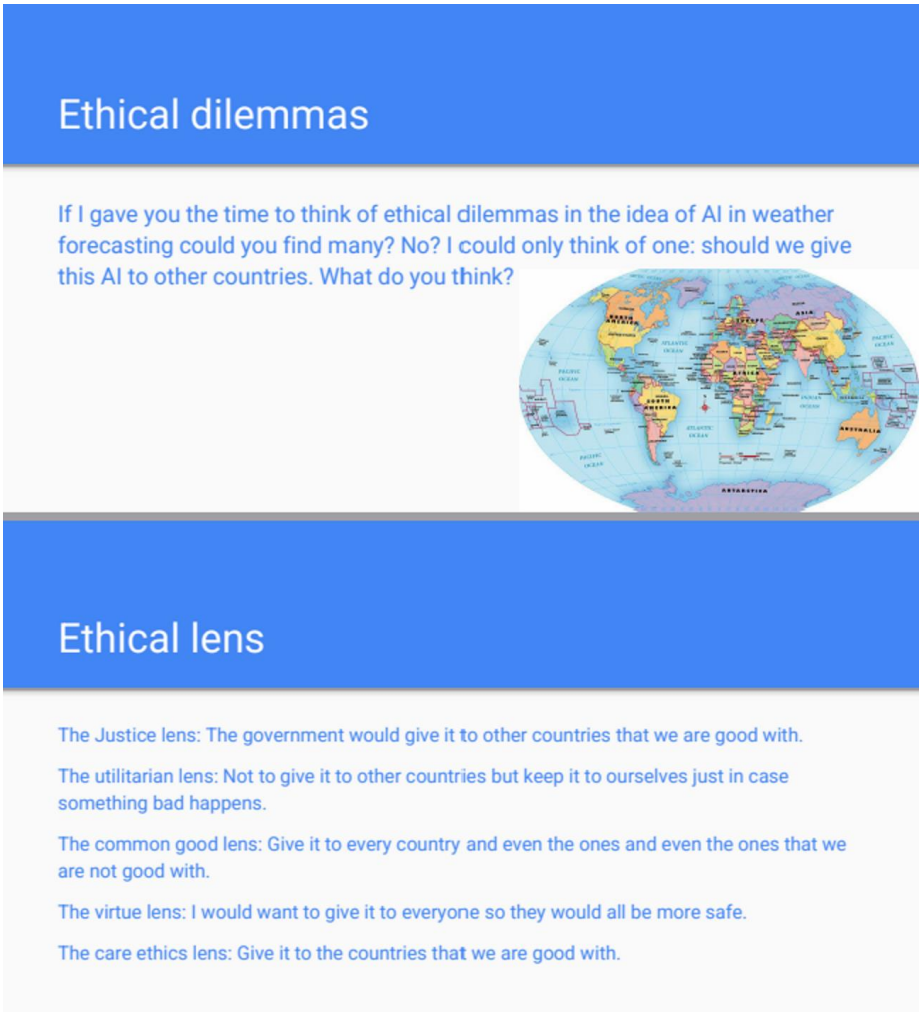
Others qualified their answers, acknowledging that knowing the theories/terms served a purpose in the academic setting. I asked Tori, “Do you think that learning the different theories is important in talking about ethics?” Tori replied,

> I guess it kind of depends, like where are you talking ethics? Where are you talking about it? Like why are you talking about it? Like when I was just talking with my sister, I don't think it mattered, but in the classroom setting, I feel like it mattered more.

I asked, in response, “Why do you think it mattered in the classroom setting more?” Tori replied, “… I feel like you can just understand more rather than having a casual conversation and it's more of, like...Like I think you should know it, but I don't feel like in order to talk about ethics, you need to know it.” I asked Dom the same question about knowing terms, and they asked, as a follow-up question, “Like as a normal person out in the world?” implying that being a student in

the classroom was different than being “out in the world.” These students saw the terminology as especially useful in the classroom setting, but discussion of ethics as something that could happen without them “out in the world.”

I asked the students to try out these normative ethical lenses on several different occasions, including, in the 8th grade class, as one part of their final projects (see Appendix D for the materials guiding this part of the project). Simon chose to complete his final project on weather forecasting AI. He had some trouble thinking of an ethical dilemma on his own or in his table group, so I helped him brainstorm. He landed on the following, dilemma, then used various lenses to think through the dilemma:



Here Simon demonstrates how using different lenses can, but not necessarily always does, yield different conclusions about what is right and could lead to different outcomes. This was a difficult task for many students, but they also seemed to value having tools to prompt and guide perspective taking, even if the application of the terminology remained imperfect. This met my goal, as I cared more about ethical theory as thought partner than as technical vocabulary, but there remains room for developing more ways to support students in applying these sometimes-strange-feeling ideas and connecting them to their own ideas about what is right and wrong.

Even when I did not ask them explicitly to use the lenses or the terminology, they brought them into our conversation. In a discussion about meta-ethics, Jimmy poses a hypothetical scenario to Sawyer, and Grey and Theo join in, drawing on the work they did on making a short video about deontological ethics (their topic of choice) earlier in the unit:

Jimmy (27:23): I have a question for [Sawyer].

Sawyer (27:23): Yes.

Grey (27:58): I just want to say that [inaudible] our whole project thing, we did a project on the deontological, and that is thinking that it's, that your attempt is more important than your [inaudible 00:28:11].

Student (28:11): Your intent, not [inaudible].

Grey (28:11): Your intent is more, matters more than. Yeah, the effect or like the outcome. The outcome. so it depends on who is judging you. Do they have a deontological stance, or whatever the other one was?

Jimmy (28:31): Non-deontological?

Sarah (28:31): Consequential.

Grey (28:38): Yeah. [inaudible]

Sarah (28:38): Okay. I have another question I want to pose to you... Do you want to have the last word?

Theo (28:47): Yeah. Okay, if we can decide if it's deontological or the other thing, that's just proving that it's man made. We can decide. Also, a bunch of things, like if, let's say, I were to

steal money, is that good or is that morally right or morally wrong? Money also [inaudible 00:29:08] the idea of whether doing something to a man made thing is good or bad is man made.

Here the students push and pull one another, drawing on their relative expertise. One way they claim their expertise is by referring to a short project that they had done earlier in the unit on deontological ethics. I had asked the students to choose a key term from our unit so far and create a "30-second story" (see Appendix D for activity instructions) explaining the term to other students (Dalton, 2021). I think this activity gave them a sense of unique insight into this theory, and I think it's notable that they chose to warrant their point by saying that they had made a video about this term.

Beyond asserting particular knowledge, the students use the ethical lenses to name different ways of determining right and wrong in a scenario of their own making, and they apply normative ethical theories and veer into meta-ethics to pose a question about moral relativity in these normative theories. They put the terms to work rather than getting hung up on getting them correct ("deontological or the other thing" gets Theo's point across, perhaps even better on second look—it does not matter what we humans label them, they are still all man-made in their opinion). In addition to providing the transcript, I also want to add how vibrant this discussion felt. The students were, on the whole, interested and invested in their own and others' ideas. The students animated the discussion, especially by proposing their own scenarios and examples to test out their ideas. Some of these student-generated test cases traveled throughout the course, too, serving as a shared reference point. In the self-driving car dilemma discussion, Hailey said, referring to a previous test example the class had discussed, that "this is like [micro]chipping babies all over again." This was a comical moment, and other students laughed, but it also showed a shared history with exploring ethical ideas (around data privacy, in this case) together.

Theme 2: Students Describe "Tech Issues Today" and Future AI Ethics Issues Differently.

In addition to providing opportunities to practice applying normative ethical lenses, I also wanted to know what students thought were the biggest ethical issues with AI, and how they might formulate their own codes of ethics. I saw this as more on the "topic" side of the topic-practice division of ethics activities, as I wanted to understand what issues were most concerning to them, and what issues they would center in their own codes of ethics. Students came up with a range of concerns, including overdependence on AI, replacing humans at work or even "taking over," privacy, malfunction, or poor design. I was surprised by how much the issue of job loss (in four out of 10 responses) came up in these responses, because it had not been a focus of mine throughout the unit.

I was curious, too, to see how these responses compared to an earlier question I had posed to them in week five of instruction: What are the biggest issues with technology today? These two different questions, at different timepoints, yielded a very different constellation of answers. Students were concerned, when asked about tech issues generally, about how humans abused technology to harm other people, mentioning cyberbullying, stalking, stealing, and hacking. June summed it up like this: "tech allows people to Bully, stalk, hack, etc." Overwhelmingly, these students wrote about tech as an enabler of humans' worst impulses. I suspect that these answers are much more directly related to the students' personal, everyday experiences with technology (especially because I asked about "today" rather than about "our future with AI"). Their responses to the first question very much mirror popular discourse around children's internet safety, which emphasizes interpersonal attacks and spreads the (often justified) fear of bad actors online. On the other hand, students' ideas about the biggest ethical issues with AI and our future focus less on individual bad actors and more on systemic disruption, like job loss. However, there are also areas of significant overlap, especially in

concerns over addiction, misuse, or over-reliance on technology. Some concerns about our future with AI are extensions of students' concerns with technology today, and some are qualitatively different visions of how humans and machines might coexist.

Just as the activity where students generated rules for their data had significant crossover with ethics, so does this indicate important things about human-machine agency. When "major issues" center poor behavior online (like students' examples of cyberbullying, toxic fandom sites, and more), students and educators may view ethics as a matter of individual choice to do right or wrong, rather than as something more entangled in both social and technological systems.

I saw a wider range of ethical issues arise out of the students' own ethical codes in their final projects. The final project had three parts that were designed to ask the students to explore an AI of their choosing through research on the technology as it currently exists (taking a technical perspective), applying ethical lenses (taking an ethical thinking perspective), and generating fiction about their topic (taking a speculative perspective). Creating a code of ethics for their chosen AI was part of the "ethics" section of the project. This activity was included in my design from the very outset—it appears as one of my design conjectures, listed in Appendix C. I theorized that creating a code of ethics would allow students to show what they thought would be right (ethical) specifically as applied to an AI of their choosing. Below is an overview of the codes they generated and ways that I have categorized the primary ethical issue(s) they addressed in these codes.

These codes touch on nearly all of the major issues I have seen raised in broader discussions of AI ethics. For example, Bea touches on the issue of AI as replacement for human labor and human accountability and control of AI; Grey emphasizes security, consent, and data privacy; Hailey looks at consent, responsibility for content moderation, bias, and the right to be forgotten; June builds a rule for continued improvement; Theo touches on the need to know

about training data and on environmental sustainability; Jimmy makes rules for art authorship and ownership; and Simon discusses open access and the responsibility for monitoring the output of the AI. I highlight these because they are such a thorough representation of the discourse surrounding AI ethics today, and I see this as a showcase of how middle school students can participate in and contribute meaningfully to AI ethics conversations; across a wide range of topics and interests relating to AI, these students created thoughtful, insightful codes of ethics.

While designing opportunities for students to be part of this conversation about AI and ethics is important, it is not the "cause" of their ability to participate, but rather the occasion; I see this design as leveraging students' own interests (through choice of topic, for example) and already-present capabilities to do deep thinking to give a formal opportunity for them to make the rules as they deem fit. Exposure to ethical theories and discussions of AI ethics (e.g., throughout the unit we talked about bias, discrimination, authorship in art, etc.) can deepen student responses, open up more informed choice for their topic, and lend a certain authority to what they write, surely, but the most crucial ingredients are already there as soon as the students arrive. However, it is also not just an individual capacity for engaging in ethical thought, or code-making in particular, that is important. Students have differing perspectives and experiences that can make ethical thinking all the more powerful when they do it together. Although the students shared their codes at the end of the unit if they chose to, we did not workshop them as a group. Incorporating collaboration through not only discussion, but also peer workshopping of individual products, will be a good next step for redesigning these kinds of activities.

Table 5

8th Graders' Codes of Ethics for the Final Project

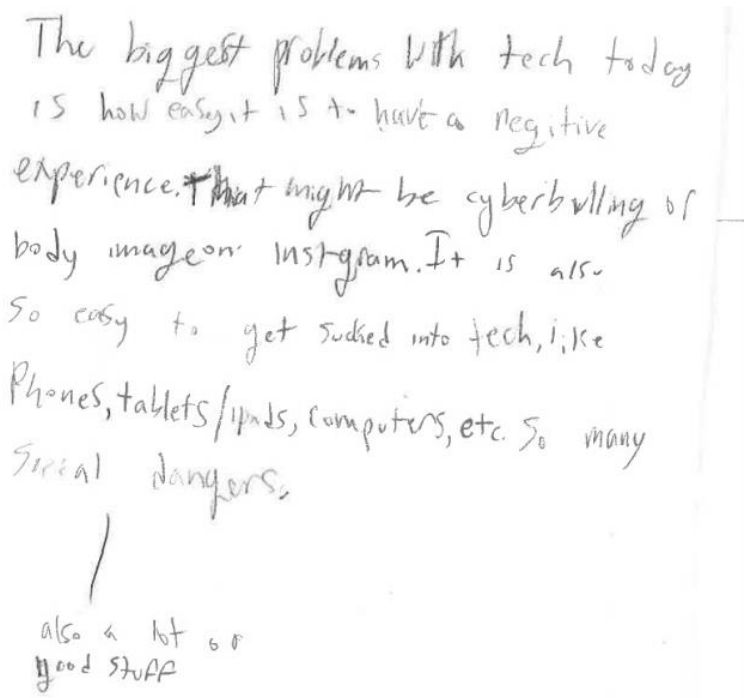
Student	Topic	Dilemma/Rules	Topic Themes

Bea	AI in graphic design	Code of Ethics: ● Keep designing AIs as helpers, not as a replacement for a job. ● Make sure the AI is under complete control of the designers. ● Pay same salary/hourly wage whether AI helped or not.	Human control over AI Replacing human jobs
Grey	AI in Netflix	Ethical Rules: Has to ask permission in order to use data Has to ask your birthday Has to ask to put a password on adult accounts	Consent Parental control Agentive AI
Hailey	AI in Social Media	Some rules to have while using social media When using someone else's data you have to have consent When asking for that consent you have to be clear on what you are going to do with thar data Everyone should be able to keep what they want to be private and to stay private There should be no AI bias in the system relating to data that was collected without authorization The responsibility on what people should and should not post is on them (still knowing they could be reported or ban) whiles till having people that will look through and take down certain post or information	Consent Privacy Bias Content moderation (human-controlled)
June	AI in advertising	Basic rules for AI in advertising: 1. AI should know when to stop digging into people's personal lives 2. Ads should not harass the potential customer 3. Information that AI found shouldn't be used to harm anyone 4. AI algorithm should always be adapting to change/new information	Agentive AI—AI limits itself Data stewardship

Theo	Underwater AI	Code of Ethics: 1. Need to know where AI's materials come from 2. Information stored in the cloud for AI needs to be held in a place with renewable source of energy 3. The government needs more resources to dedicate to environmental control	Materiality Environmental stewardship
Jimmy	AI Art	3 rules for art AI if it was used: May only be used as a reference for artists. AI that is open to the public may not be used to make a full piece of art, only a sketch Framing someone for a crime with hyper realistic generated art is a felony	Authorship Deepfaking
Simon	Weather Forecasting AI	1. Make sure to be constantly overwatching it to make sure nothing bad happens. 2.Make sure to give it public use for everyone to use. 3. Let people read it even if they don't work there.	Human control over AI—AI stewardship Equitable access

Theme 3. Persistent Enthusiasm and Skepticism, but No Consensus around "Positive Ethics" for AI.

Throughout the unit, I incorporated opportunities for students to think about the positive current and future contributions of AI might be. This was primarily borne out of a fear that I would leave students scared and defeated by discussing many of the alarming ways that AI does and could cause harm. However, students also showed me their genuine enthusiasm for and belief that AI would be beneficial for humanity. This was especially evident in the case of Bea, an 8th grader. Bea consistently resisted ways that I (sometimes inadvertently) framed technology and/or AI in a negative light. For example, in responding to my question, discussed in more detail above, about tech issues today, Bea made sure to add in that she saw "also a lot of good stuff":



I can still picture the joy and excitement on her face when she did her final interview, telling me about how cool Sophia, the first AI to gain citizenship, was, and how she was excited about machine sentience:

Sarah (03:18): Have any of your views about what AI is or where it is or how it's used changed since the beginning?

Participant (03:19):Well, where it is and how it's used has changed but my opinion is still pretty solid.

Sarah (03:19):What's your opinion?

Participant (03:19): I love AI because it's really cool and I think Sophia is so cool.

Sarah (03:19): Yes. The first AI to get citizenship? Yeah, that's a really interesting story. Um, yeah. Do you think there could be more AIs in the future that would be-

Participant (03:19): Definitely. I don't see how there wouldn't be.

Sarah (03:19): And is that, does that worry you? [Participant shakes head no] or no? Why not?

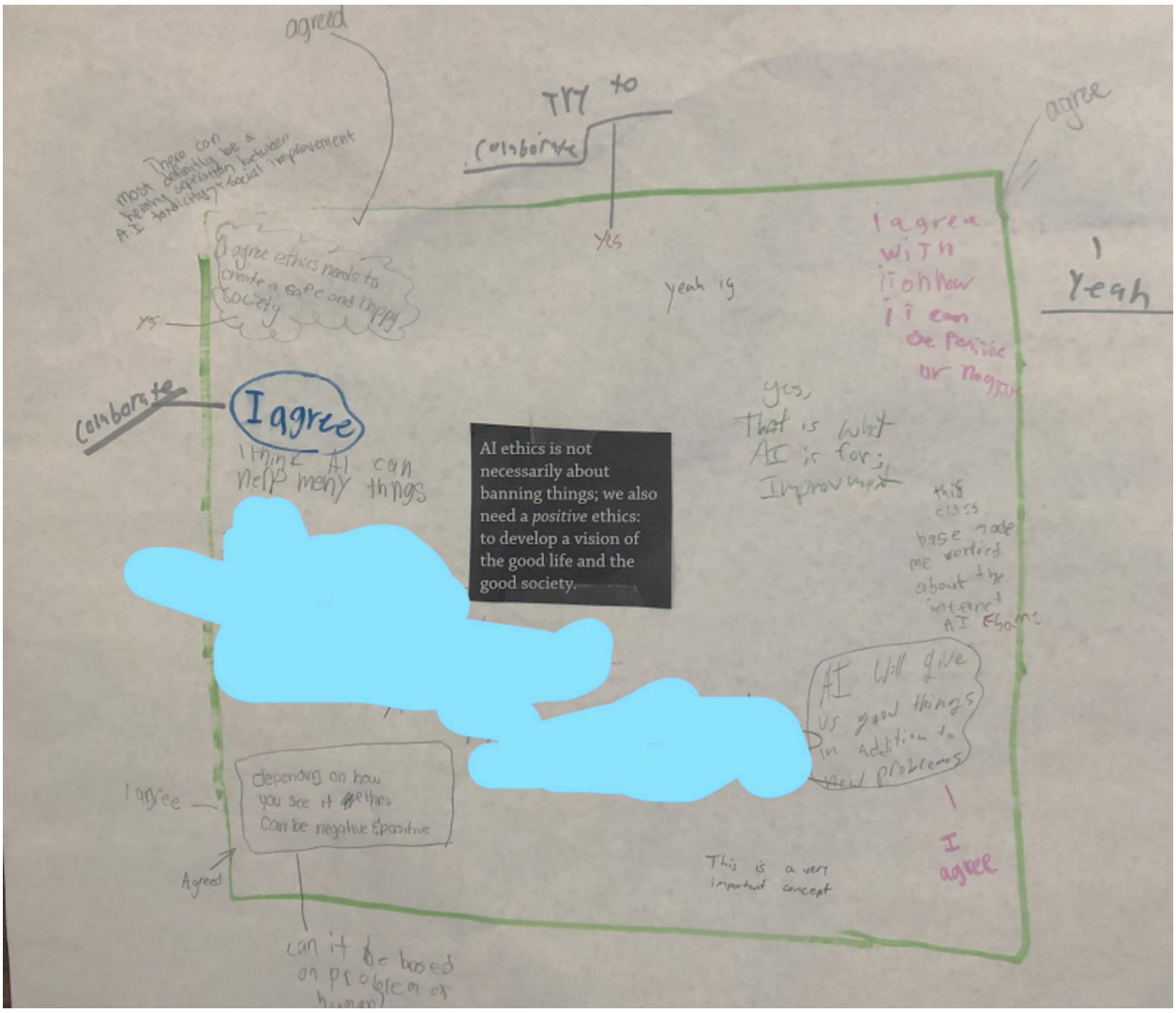
> Participant (04:28): Because well I think we're gonna be better about stuff in the future. [big sigh. Pause] If they gained sentience, that's even cooler because that's like, "Oh that's AI but it's, it has feelings." That's amazing. I don't see it as scary. I think it's really cool.

To Bea, machine sentience is “cool” and “amazing” rather than scary, and she has faith that “‘we’re gonna be better about stuff in the future.” Bea is but one example of students’ persistent hope in future AI and belief in its power for doing good in the world. This hope and excitement appeared again and again throughout the unit; the students were frequently awed by AI’s abilities and enthusiastic about its potential. I think this is a beautiful thing, but I wonder what it means about and for critical instruction around AI and ethics. Surely, I do not want to crush a student’s excitement, but I do want to make sure that this excitement does not also blind them to current and potential harms of AI.

On the other hand, Hailey, Simon, and Sawyer were persistently skeptical of AI’s role in our lives. From the very beginning and to the very end, these students remained stalwart in their view of AI as overwhelmingly harmful. These cases of persistent feelings about AI on the extremes indicate that, while their knowledge of AI certainly changed, their attitudes about its ethical impact did not. This warrants further exploration in future studies and designs, as it appears that firm stances toward AI as positive or negative followed these students through almost every activity (e.g., pre-writes, discussions, final projects). Does this foreclose consideration of other perspectives, and/or does it indicate passionate, well-thought-out conviction?

I think that Bea felt she was resisting negative portrayals of AI throughout the unit. Focusing on ethical dilemmas can surely make ethics seem like it is *only* about problems. I tried to counter this by discussing “positive ethics” and asking students to consider the contributions and benefits of AI. Near the end of the unit, we did a “discussion” through writing, where students responded to prompts from an AI ethics text. Once everyone had a chance to respond to the original text, I asked them to go back around, read others’ contributions, and respond to

the first round of comments. The chart paper below (with unconsented student contributions redacted) shows this discussion with the seed quote/original text, “AI ethics is not just about banning things; we also need a positive ethics: to develop a vision of the good life and the good society” (Coeckelbergh, 2020, p. 175).



I designed this activity to complement our whole-group and small-group discussions, giving another way for students to participate and respond to one another outside of the structure of normal “speaking turns.” I also chose this particular quote as one of the focal texts because I wanted to counter the many opportunities we had in class—and, moreover, that students are exposed to by virtue of deficit-based ideas about and punitive impulses toward

young people's technology use—to think about "dilemmas" and "issues" in relation to ethics. Some students wrote about how the good and bad come together, or are a matter of one's viewpoint: "AI will give us good things in addition to new problems" and "Depending on how you see it ethics can be negative & positive." Overall, these students expressed—here and in other places throughout the unit—high hopes for AI to "help," "collaborate," and "improve," but they also recognized that there is no such thing as a purely positive ethics; there's always a new problem or another perspective "depending on how you see it." However, one student wrote, "This class [has] made me worried about the internet AI," and another student drew an arrow and wrote "same." This made me wonder if I had designed enough of this "positive ethics" into the course, where we would think together about our vision for a good life.

Bea's insistence that AI—and people—would be good, along with students' written comments that the class made them worried, made me realize that, though it was a goal of the curriculum, I did not do nearly enough teaching around a "positive ethics" for AI. By the end of the unit, we did not reach a consensus about what this "positive ethics" could look like, or even why it might be desirable.

Theme 4: Peers as Ethical Thinking Collaborators in Understanding Meta-Ethics: Sharing and Challenging Others' Perspectives in Whole-Group Discussion.

Students had several opportunities to discuss different ethical perspectives as a class, including in response to a self-driving car dilemma video (Lin, 2015), in a group discussion about different types of meta-ethics where we thought about whether there might be a moral truth "out there" to discover, or if it is all a human invention, and in a fishbowl-style discussion after watching the science fiction film *Ron's Gone Wrong* (Smith & Vine, 2021).

In these discussions, group dynamics and student personalities and personas played an outsized role. For example, Sawyer took it upon himself to play the antagonist, challenging his classmates in both productive and disruptive ways. When asked if they encountered any views

that challenged their own, two students independently named Sawyer. Sawyer played an important role as a foil but also could foreclose others' participation by dominating (in volume, frequency, and fervor) the conversations. In the same discussion on meta-ethics where Jimmy challenged Sawyer, posing an ethical dilemma for him, Sawyer tried to offer a definitive answer to whether we can program morals into machines. However, once other students offered complicating opinions, Sawyer replied but decided he needed to think about it more. This could indicate that in addition to challenging his classmates, his classmates also challenged Sawyer to go beyond his initial reaction and impulse to take a firm stand. In addition to seeing these group dynamics unfold, the conversation is also interesting for the sophisticated ideas that the students try out:

Sarah (32:49): Okay, bringing it back to AI and ethics: Do you think, given our conversation about meta-ethics and your own ideas about whether or not ethics exist, are made up by humans, or exist outside of them, do you think we can create a machine that has morals?

Sawyer (33:11): No, not at all.

Sarah (33:13): That was quick. [Theo]?

Theo (33:14): Because it's…So, I personally think it's man-made thing. So yes, technically, we could program it to have morals, but then again, everyone has different ethics and morals because there isn't any one thing, like there isn't a set of rules saying, "This is right. This is right. This is wrong. This is wrong," and what people think can always change based on what they observe. Like if it's to a car situation and you know, like "Oh, I know that my parent got hurt in this situation if like the car would have went that way, so I'm going to go this way because of that," but originally they thought, "Oh, you should go this way or not [inaudible].

Sarah (33:57): So what do we do if we can't program et—morals into a machine? Yeah, [Sawyer]?

Sawyer (34:03): So, it depends on the morals... Thing is, maybe it could be good for that time, but in the future, let's say thousands of years in the future, things will be different. There would be maybe different morals. Like if we went back maybe 500 or 10,000 years ago, people thought it was okay to kill each other. But now, we see it as something that's not, that's not great in our somewhat morals. Um. So, if we had one person do that, he would just put what he sees as moral, or even a group. Maybe, let's say, thousands of people put in these morals. It will still not really, really... Have to think on this more [inaudible 00:34:51].

The conversation continues, with other students stepping in to offer their ideas once Sawyer has stepped back. The students (and I) discuss what might happen if you tried to

program an AI to have morals. I ask what might happen if the data set used to program this AI is human history:

> Tori (36:10): I think that um if you put in history into an AI to give it morals, then it will just have the same progression in morals as we have, from what we've learned, history, or from personal experience.
>
> Speaker 13 (36:33): [inaudible 00:36:33].
>
> Tori (36:33): Yeah. So, that's the thing is that it couldn't catch up to us unless we reprogrammed it with the newer history that's happening every day.
>
> Sarah (36:37): And this is something that people try and do is say, "Okay, wait, I programmed this AI and I fed it a bunch of bad data," because a lot of our data from history is racist, sexist, um really terrible in a lot of ways. And so, if you use that historical data to train your AI, you're going to get a biased AI. And then so, some people say like, "Okay, well, what I need to do to fix this is to change how the AI works, to change how it's weighting different factors," um and maybe, some people believe that they can make an AI that makes better decisions than people, and other people strongly disagree with that. Yeah?
>
> Jimmy (37:18): I think a lot of these AIs that we're like putting a lot of these situations in are like really just stupid and unnecessary. And like I think that it's... You have to make a decision with a machine that is deciding who should live and who should die but that machine is completely unnecessary [inaudible 00:37:51]…hands, and not in a good way.

Here the students and I make connections between ethical principles and technical capabilities, thinking about how an AI's dependence upon particular data can shape its ethical capacities. Tori points out that, if an AI uses historical data, it not only will reproduce our human biases but will always be behind us, playing catch up to our morals. This was the kind of interaction I had hoped for in my design, where students blend speculation and technical knowledge to generate critical insights.

Jimmy also points out that we (humans) might be needlessly putting AIs into situations where they make life or death decisions. Jimmy brings us back to that most basic of AI ethics questions: Should we even make or use it? In this discussion, I saw several students offer up big ideas and take their classmates' big ideas seriously too. I think this conversation was so rich because the students had the ingredients for productive dialog: their peers, with diverse interests and perspectives, who were listening and responding, some lenses and language to

think with/question/challenge, permission to explore ideas, even taboo ones, and an interesting topic.

There are both designed and “undesigned” factors that made this a rich discussion. Watching and discussing the film came near the end of the unit, and I can see that students incorporated some of the things we previously discussed in the unit (like meta-ethics and the impact of training data). There are also “undesigned” factors that made it especially interesting, like the fact that the school had very recently implemented a phone use policy that had clear connections to the way B-Bots were regulated in the film universe. Although this undesigned factor proved to be fodder for an interesting discussion (see Example 1 in the next chapter for more on this), setting up and placing the film as I did came with certain affordances for rich ethical discussion. Giving the students a framing question for the film—What commentary is this film making about society and technology?—was helpful in cueing the students to attend to the thorny sociotechnical issues raised in the film, which in turn prepared them for the discussion. In addition, I asked them to gather their thoughts on the QCQC graphic organizer (see Appendix D), which they could draw from (and literally bring with them) into the fishbowl circle. This combination of previous conceptual exposure, framing, and supportive documentation, plus things less under the designer’s control, like timeliness/direct relevance to a real-world issue and extant group dynamics, all came together to make a meaningful discussion.

Thinking with Theory and Data for Redesign around Ethical Thinking

This approach to ethics was heavily reliant on normative and dominant perspectives, so (re)thinking with posthumanist theory here is especially important. Posthumanist ethics would exhort us to include nonhumans in our notions of what is right and wrong. This is not any sort of favor that humans would do for nonhumans, but rather a necessity, and a reflection of the way we are bound together. It is not generous inclusivity that motivates this move, but rather the recognition that there is no human without these other things. Taking the example of Theo’s

code of ethics for their underwater AI, the first rule is that you would "need to know where AI's materials come from" and the second states that "Information stored in the cloud for AI needs to be held in a place with renewable source of energy." This is a beautiful expression of posthumanist ethics, where the right thing to do is not merely about human behavior, but about acknowledging our interdependence with sources of energy and physical materials. It reminds me, too, of Bennett's (2010) discussion of ethics (and, relatedly, agency). She writes,

> Perhaps the ethical responsibility of an individual human now resides in one's response to the assemblages in which one finds oneself participating: Do I attempt to extricate myself from assemblages whose trajectory is likely to do harm? Do I enter into the proximity of assemblages whose conglomerate effectivity tends toward the enactment of nobler ends? Agency is, I believe, distributed across a mosaic, but it is also possible to say something about the kind of striving that may be exercised by a human within an assemblage. (p. 37-38)

Theo's idea of ethics encourages us (humans) to seek out an ethical assemblage—where you know the provenance of AI's materials—and also to make them so by choosing a server location powered by renewable energy. This differs from Bennett's (2010) idea of ethical participation in a key way, though: Bennett downplays the effect of individual action to an extreme degree, focusing instead on how assemblages, rather than individuals, have trajectories that can harm or "[enact] nobler ends," and our ethics then must relate to opting in or out of these assemblages. But I wonder, Is our human agency so constrained as to make ethics distillable to this single choice? I think designing curriculum around this view of ethics is practically undesirable, even if it could possibly be "true." Such a strong negation of individual human agency might foreclose imagination around what is right, and I still think that hope should be protected at all costs (even if it is rationally "incorrect"). However, the idea of the

(more or less) ethical assemblage could be a powerful one to include in future instruction, and I would like to include it as an option when introducing the various ethical lenses.

Section Conclusion

My goal was never to be dogmatic about ethical vocabulary, but I hoped that learning about the different (dominant, Western) ethical theories might give students some ways to think outside of—and question the underpinnings of—their own perspectives. Although I did not want to overemphasize the terminology or foreclose other (nondominant) ideas, I felt that using these terms (carefully) was ultimately warranted and even rewarding. It felt to me like learning the theories—then questioning what principles they take for granted by discussing meta-ethics—gave the students license to explore ideas and opinions that otherwise might have been taboo, especially in the classroom.

Perhaps more important than this exposure to any particular content was exposure to others' ideas and opinions. When I asked "and do you think it's important to have that exposure to different people's opinions in forming your own?" in her final interview, June sagely said, "Yeah. I just think that's a great part of education and growing up and learning." This is a case of design being somewhat at the mercy of context; we might not choose which students are in our class like we can choose our activity structures and materials. However, we can and should be responsive to them, making space for students to express existing views and try out new ones with each other.

Chapter 6: Fiction and Speculative Thinking for Critical Posthumanist Literacy Pedagogy

The three big questions guiding Chapters 4 and 5 each center a specific concept that was included in the design (ontology, agency, and ethics/justice) of the curriculum. This chapter focuses more explicitly on the "how" rather than "what" questions that animated the project: How should/could we teach about AI and ethics? Specifically, this chapter takes up the final project goal listed in Chapter 3: using the consumption, production, and discussion of science fiction/speculation as a pedagogical tool alongside technical activities that demonstrate core aspects of AI's current functionality.

In this section, I draw on data to begin answering the following research questions:

- How can I incorporate science fiction reading/viewing and writing, alongside opportunities for technical engagement with AI, to foster agentive participation and ethical engagement around AI?
- (How) might approaching AI ethics from a speculative/imaginative perspective strengthen student engagement with and understanding of AI?

The focal data for this section include:

- Transcripts from our class discussion of a science fiction film
- Students' completed graphic organizers responding to the film
- Speculative fiction writing samples
- Responses to a pre-writing question about fictional AIs
- The speculative fiction-writing portion of 8th grade final projects

One goal of this project was to expose students to AI in ways that were not dependent upon computer science expertise and equipment. For me, as a former librarian and literacy researcher, science fiction was an appealing pedagogical tool to use toward this aim. This is not a new or original idea; for example, the MIT AI and Ethics middle grade curriculum also incorporates speculative fiction writing as a teaching tool in their AI ethics unit for middle grade

students (Payne, 2019). Beyond this specific example, fiction has been used as a (formal and informal) teaching tool for as long as people have been telling stories. However, fiction consumption and production remains a pedagogical approach that is less familiar to computer science curricula than programming exercises. The critical posthumanist perspective can be a valuable one for guiding the design and incorporation of fiction into AI ethics curricula.

Of course, this is also another way to read and write with machines, not as the medium of composition or production, but as the (imagined) subject/object/character. By writing and reading about or as (in the case of those who chose to write from an AI's perspective) an AI, students gain yet another perspective on what it means to do critical posthumanist literacy.

In this chapter, I explore using fiction in critical posthumanist literacy pedagogy. I have included an example of producing and consuming science fiction texts, as they are both valuable options for instruction with different affordances for students. Most importantly, the act of creating or reading/watching fiction is, in itself, necessary but not sufficient for cultivating criticality and performing ethical thought experiments.

As I analyzed the instances where students engaged with science fiction, I realized that it was serving several distinct functions: as a foil for thinking about current technology and "real life," as a space for ethical experimentation, and as a way to prepare for or practice responding to future technology. In the sections below, I also explore the other factors that are essential for making science fiction pedagogy joyful, impactful, and effective: peer dialogue, supportive structures, and rich texts. In each section, I use an example activity (or two) to illustrate a function of using science fiction in the classroom and a specific supportive technique for doing so. Finally, I conclude with students' and my own reflections on the benefits and drawbacks of using fictional AI to help us understand AI more broadly.

Selected Examples

Example 1. Using Discussion to Find Echoes of Real Life in Ron's Gone Wrong

In this section, I will highlight one function I saw science fiction pedagogy serve—as a foil for thinking about real-life AI—and two techniques for using science fiction in the classroom—(loosely) structuring dialogue and providing a space for students to organize their thoughts beforehand.

One way to use speculative fiction in teaching is to watch science fiction films and discuss them. That is what we did in both the 7th and 8th grade classes, screening the animated film *Ron's Gone Wrong* (Smith & Vine, 2021). In *Ron's Gone Wrong*, Barney is a middle school student whose family cannot afford the companion robot ("B-Bot") that everyone else has and that dominates middle school social life. Barney's dad and grandmother manage to secure a B-Bot that has fallen off a truck, giving it to Barney for his birthday. However, the fall from the truck has damaged the bot, and instead of being Barney's "best friend out of the box," Barney must train the bot, Ron, to be his friend, all while keeping him from being destroyed by the B-Bot making corporation for being defective and lacking the normal safety controls of the other B-Bots. The film touches on issues of data privacy and children, the power of corporate tech, and the meaning of authentic friendship.

Before watching the movie, I primed students to be thinking about what commentary the movie was making about our lives with technology and asked them at some point to jot down an important quote while they watched (and ate popcorn and snacks). After watching the movie, I asked them to fill out a "QCQC" (Quote-Comment-Question-Connection) Sheet I had borrowed from Dr. Hundley's YA Literature course (Hundley, 2021) to organize their thoughts for our fish-bowl-style discussion. For the discussion, a group of students volunteered to sit in a circle while the rest of the class observed around them, listening to their discussion. When a student in the

outer circle wanted to contribute or felt that a student in the circle had contributed enough, they tapped a seated student to take their place.

The students immediately saw connections to the real world, postulating that the company in the film was supposed to be Apple, and the money-hungry, black turtlenecked CEO was meant to be Steve Jobs. They spent some time debating the plausibility of the plot, portrayal of parents, and even the cost of the robots. They also connected the treatment of the B-Bots in the movie to the phone policy that had recently been implemented at their own school. The parallels were striking: for example, in the movie, each student had a shelf at school where they put their robot at the beginning of the day, just as the school had a slot for each student to leave their phone upon entry. In the following transcript of the 8th grade fishbowl discussion, Sawyer and Mia liken the B-Bots to phones and the surrounding rules to the ones the school has around their own phones.

Sawyer (27:00):Why would they, first of all, allow B Bots at school? I know they blocked it but why would they allow B Bots in the first place? And how do they help you learn how, why do they have, like, little things? It just doesn't really make sense. They should at least block it. Like, you know, they have rules about phones, which is technically “AI” or “robot”[makes air quotes], so around here you're not allowed to bring your phone in class. What makes phones any different from these B Bots things?

Grey (27:32): They brought them to school because a lot of people were using them to get to school. They used them for communication, so that's another reason. And they had [inaudible 00:27:41] but they weren’t allowed to bring them. And also, one kid was using it for a presentation. So that's also [inaudible 00:27:46]

June (27:46): So basically, phones are different from B-Bots because—the B-Bots were like mostly [inaudible—a kind of?] of AI and phones are mostly just computers, which…yeah.

Mia (28:34):I think that they are similar to how robots do it. They might bring the B-Bot to school and use it occasionally, but kind of how we have our phone protocols here, you put it on a shelf and don’t use it for most of the day. They do, they do kind of... they don't use it until recess anyway. And I hear from what [June] said, I think there's a lot more [inaudible 00:28:40] there are similarities, but there are also a lot of differences.

Here the students work to make distinctions between the functions of B-Bots and their own phones—are they AI? Are they just like B-Bots?—in order to sort out what they thought about the rules surrounding B-Bots (and I later asked about the rules surrounding their phones).

They also discussed the ethical dilemma of high cost to access the technology in the movie and the inequities this created, and whether robots could or should be considered “friends” for humans. They wondered whether it was realistic that so many students could afford the B-Bot, linking their understanding of technology access in their world to that of the film.

Furthermore, watching and discussing the film in the context of the unit and with peers revealed new ideas. Bea had watched it two times before, but made new connections this time. In her final interview, Bea and I discussed her experience with the film:

> Participant (09:44): I've seen it three times now. I love the movie. I thought that it was very good and I love how it talks about AI.
>
> Sarah (09:55): Yeah. Did you see anything new when you watched it and discussed it with your peers?
>
> Participant (10:01): I realized how much AI... My brain didn't connect that it was about AI. It just was like, "Oh, there's this boy Ron, this b-bot that can think now. That's pretty cool". Not that it was machine-learning or AI. That's machine-learning, right?
>
> Sarah (10:27): Yeah, so it's fictionalized but what it uses to take in all that data and knowing things about the user, I would say that's it. [Participant: Okay] Although I don't know what like, because it's fictionalized, I don't know exactly what the details of the machine-learning look like.
>
> Participant (10:34): Well, he was learning. Like with the sticky notes like he replaced some [inaudible].
>
> Sarah (10:58): Yeah. I think we can use it to understand things about machine-learning too. The technical details I don't know about.

Here I want to point out how this data excerpt demonstrates both the power and limitations of using a science fiction film to think about our lives with technology. Bea sees something new and different when she watches and discusses the film with her classmates. She connects the technology in the movie to a current technology, wondering if it really fits in the category of machine learning, and is curious about—then generates a theory about (mentioning how the B-Bot learns in a scene where he studies sticky notes)—the technical operation of the technology portrayed in the movie.

However, there is a point at which the conversation stalls because we do not have access to details about how the fictional technology is supposed to work. In fiction, the author/creator gets to imagine and invent technology that might reflect some—but not all—aspects of our current "real" technology. This both allows us to ignore technical functioning in important, productive ways—allowing us to think through ethical situations without needing to know how to code or train an AI model—but also can hinder our ability to foresee some ethical risks. For example, there are ways that these particular B-Bots are designed that beg us to think about how children are being surveilled in the fictional world, and how we might want to be careful about that when we place surveillance devices (e.g., Amazon's Alexas) near kids. There are also ways that the technical functioning of the B-Bot might obscure its connection to technology in use today, allowing people to dismiss it as purely fiction because we don't have the current capability to create a B-Bot that looks and functions like it does in the movie.

Example 2. A Cautionary Tale: June's Science Fiction Story as An Ethical Thought Experiment

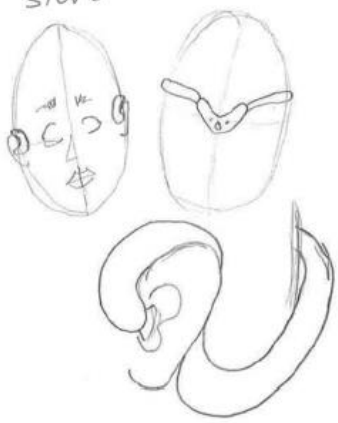
In this section, I highlight another function and supportive technique: fiction story as ethical experimentation and developing writing supports. As I wrote in the previous chapter, one way we "built an AI" was by inventing one and putting it into a story. To support students in their writing, I created a guide for the activity that included questions for the students to consider when coming up with their AI and story surrounding it. The sheet looked like this, below, which also shows 8th grader June's responses:

Build Your Own AI: Sci Fi Edition

Characteristics of your AI:

- What can it do? Block People in/mute words + sands
- What does it look like? (Draw it if you'd like!)
- Who made it, and why? To tune out
- What problem(s) does it (try to) solve?
- How does it work? Process es words and mutes
- What kind of data does it need? Sounds
- Who uses it?

slurs

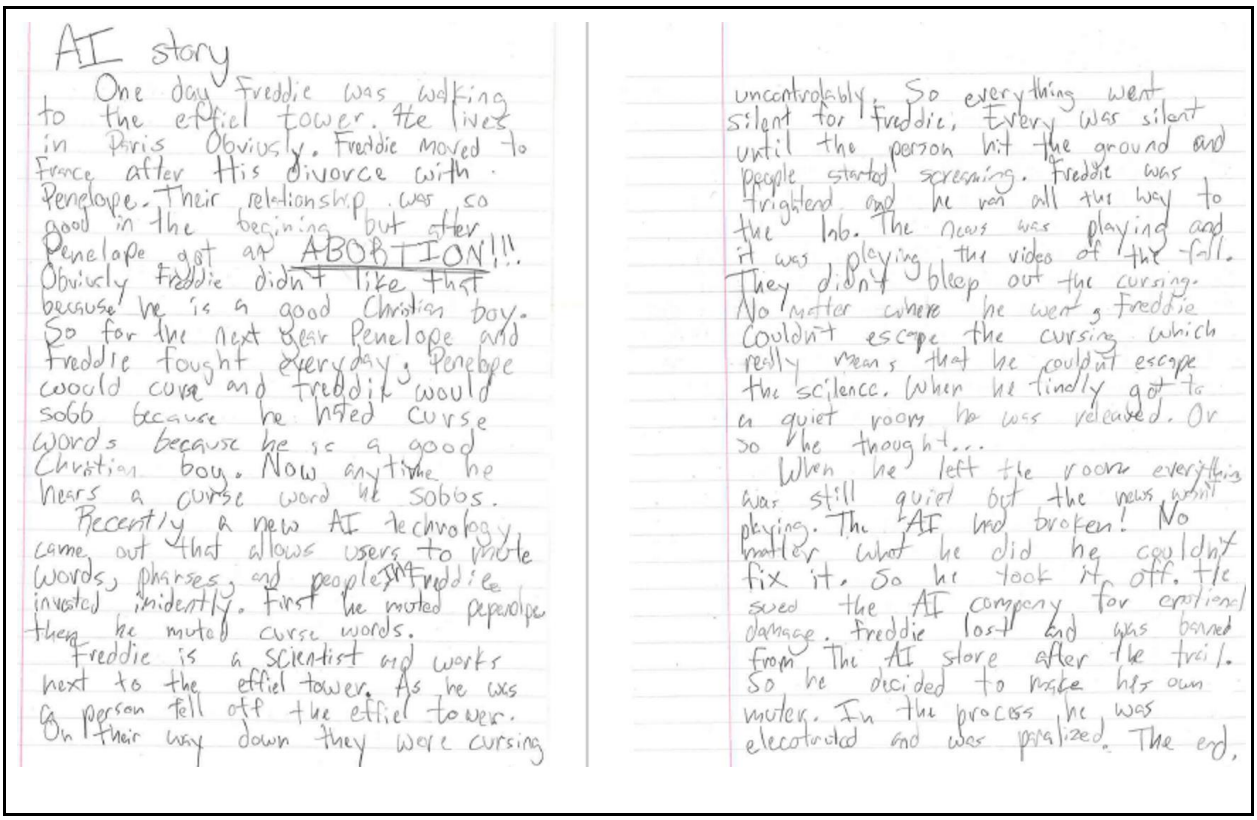


Putting your AI into a story:

- Now let's put your AI into a story.
- Pick a
  - Setting (place and time)
  - Point of view (who's telling the story?)
- Make sure to include some
  - Dialogue (what characters say)
  - Imagery (visually descriptive language)
  - Characters
- Plan a plot (what happens?)
  - Create a conflict
    - What dilemma does your character face?
    - What options do they have for reacting to it?
    - How do they resolve it?

Setting:
P.O.V.: ~~Freddie~~
Characters: Freddie + Todd
Problem: The AI breaks and mutes everything

June went on to write a poignant and darkly funny moral tale about her fictional AI creation, a device that could mute particular people, words, and sounds. Below, I have included June's final story.



AI story

One day Freddie was walking to the effiel tower. He lives in Paris Obviusly. Freddie moved to France after His divorce with Penelope. Their relationship was so good in the begining but after Penelope got an ABORTION!!! Obviusly Freddie didnt like that because he is a good Christian boy. So for the next year Penelope and Freddie fought everyday, Penelope would curse and freddie would sobb because he hted curse words because he is a good Christian boy. Now anytime he hears a curse word he sobbs.

Recently a new AI technology came out that allows users to mute words, phrases, and people. Freddie invested indently. First he muted peperdipe then he muted curse words.

Freddie is a scientist and works next to the effiel tower. As he was a person fell off the effiel tower. On their way down they were cursing uncontrolably. So everything went silent for Freddie. Every was silent until the person hit the ground and people started screaming. Freddie was frightend and he ran all the way to the lab. The news was playing and it was playing the video of the fall. They didnt bleep out the cursing. No matter where he went, Freddie couldnt escape the cursing which really means that he couldnt escape the scilence. When he finally got to a quiet room he was relaxed. Or so he thought...

When he left the room everything was still quiet but the news was playing. The AI had broken! No matter what he did he couldnt fix it. So he took it off. He sued the AI company for emotional damage. Freddie lost and was banned from the AI store after the trial. So he decided to make his own muter. In the process he was electrocuted and was paralized. The end.

This activity allowed June to experiment with, and comment on, ethics surrounding her invented technology by creating a fictional world, including a main character with a particular set of ethical values. I want to point out a few things about the world June creates in this rich text: first, June explores several options for taking action in the face of corporate AI, including opting in for convenience or the (misguided) dream of improving quality of life, and (the impossibility of) opting out safely. Second, June remixes current ethical issues with futuristic ones, reflecting some of current discourse around abortion and what it means to be “a good Christian boy” in her speculative story. She paints the picture of a highly litigious society where the company triumphs over—and destroys—the consumer legally and then physically. (We do not know what becomes of Penelope in this story.) This story is a nightmare of being destroyed by a product purchased to marginally improve quality of life or cater to some egoistic whim. There is no redemption for Freddie’s follies—first in shaming and silencing Penelope, then in trying to take on the AI company, then in trying to develop his own muter.

In this story, we can learn a lot about June's thinking about the comparative risks and benefits of the technology she has imagined: on one hand, we have a product whose major benefit is that it might make certain (privileged) people more comfortable by protecting them from speech they find disagreeable; but on the other hand, the risks include everything up to untimely death by electrocution. To me, this imbalance is an exaggerated version of what happens when students choose to use many AI-driven tools today, choosing personal convenience despite risks of grave harm.

This story is a cautionary tale, with Freddie being the embodiment of shallow—and punishable—ethics. Although he is nominally pious, he mistreats Penelope and seeks a convenient, tech-enabled way to silence her. In June's fictional world, Freddie makes a poor risk-benefit calculation about the technology, embracing it as an easy fix for a "problem" that probably would have been better addressed by other means. June's story is an ethical experiment that asks and answers the question, "What might happen if people could use technology to block out things they find objectionable?" As her teacher, I get a glimpse into June's thinking about this ethical dilemma and am drawn into the fictional world where she answers the question she has posed.

In the planning document, I asked students to imagine how the device functions, in an effort to draw in technical details (currently possible or imagined) into their stories. June's invention is an AI-driven piece of hardware, drawn on the sheet above, that is designed to selectively mute certain sounds/people. However, the device's *mal*function rather than its function becomes the crux of the story. June here touches on two ways that students expressed concern about AIs: when they function as designed, but are designed poorly, and when they break. These are two different sorts of threat models with breakdowns at different stages in the product's life cycle. I argue that fiction stories like these allow us to see both, in the span of a

few short paragraphs, the intended and unintended effects of a technology like June's, and others.

Example 3. Envisioning the Future: Writing Tiny Dystopian and Utopian Fictions about AI

In looking back at utopian and dystopian student fiction-writing during my analysis, I realized that this activity might serve as practice or preparation for things that could or should happen in the future. In this section, I discuss the speculative pedagogy function of practicing/preparing for the future, along with the technique of designing speculative prompts.

The final piece of the 8th graders' final projects was to do some speculative fiction writing and invent an "artifact" from the future. I asked them to write a mini-story imagining the best and worst possible future worlds with their technology, giving them brief prompts: ""It's the year ____ and things are not looking good" and ""It's the year ____ and things couldn't be better." I described the rest of their task this way:

> Now choose a scenario (your worst-case or best-case from above, or a blend of elements of both/something in-between!) and create and describe an artifact that relates to your technology from your future society. This could be an ad for a product, a piece of art, a piece of writing (a social media post, a news article) from your future society. Write a brief description of your artifact. The description should provide context for the reader to understand the artifact—When/where did it come from? Who made it? What does it mean? For example, you can imagine that the description is what would accompany the artifact in a museum or that the description is what someone would have used to explain why this artifact was put into a time capsule.

I designed this set of prompts to give the students two ways to envision the future with their technology: through creating an "artifact" and by writing tiny utopian and dystopian stories (the latter of which is similar to Payne's (2019) speculative fiction activity). The former allowed them to leverage their multimodal creativity to place themselves into their imagined world. The

latter was designed to be a low-entry-barrier way to explore ethical risks and benefits of the technology.

Below I have included a particularly prescient example, Jimmy's imaginings about the future with AI art. Keep in mind that Jimmy wrote this after we discussed DALL-E, the image generating AI, in class. In brainstorming before writing, Jimmy made the following notes about how they thought their AI might be used in the future:

> 2122:
>
> I think a lot of AI art will be used in a hundred years, but maybe not in the sense of actually using it as the end product art. I imagine that artists will use deepfakes and AI generated art as a reference picture. I could see where that could be dangerous for videos and faces, so I bet those will be banned. I think there will be more in depth deepfakes/AI generated art that will be advanced and safe enough for the general public to use.

I then asked the students to "Imagine how things could go wrong in the future with your AI. Could using (or not using) your AI lead to suffering or injustice? Describe a dystopian future here" and provided them with a short writing starter if they wanted to use it ("It's the year ____ and things are not looking good."). Jimmy wrote:

> It's the year 2122 and things are not looking good.
>
> Programmers have made a system where if you can give them enough money, they will make a realistic video of anyone doing anything. There is no way to tell what is real and what is fake. Court cases have been made unfair with fake footage and politicians have been "filmed" saying awful things.

I also asked them to "Imagine how your AI might be used to make the world a better place. How might your AI benefit humanity/the planet/other life forms? Describe your best-case,

utopian future here" and gave them the utopian version of the story-starter ("It's the year ____ and things couldn't be better."). Jimmy wrote:

> It's the year 2122 and things couldn't be better.
>
> Artist communities are thriving because of the new ability to generate a reference photo with keywords, and art is now something anyone can make!

These little bursts of fiction gave me as an instructor and scholar such rich information about how the students were thinking about their chosen technology. I could see what was worrying the students, and what they were hopeful or excited about. I (and they) could look at their utopian and dystopian scenarios side by side and consider how the risks and benefits compared. This could be a wonderful starting point for further discussion, too: looking at your scenarios, do you think the promise of this technology is worth the potential risk? I hypothesize that engaging in this kind of thinking and creating is good preparation, good practice, and good habit-building for having critical discussions about technology in our daily lives.

Jimmy's artifact is a blog entry from 2035, and it reads: "A programmer by the name of Jimmy Smith has created a new program that allows people with no artistic ability to create art. Jimmy says it's simple to use: Input a number of keywords between 1 and 15 and the AI will incorporate these words into a drawing. Teens on the app ClikClok have already begun a trend where they enter their name and other various personal information to see what art it makes. However, there has been backlash from artist communities who claim this AI will put them out of business. Jimmy Smith is yet to respond to this accusation, more on our main page as (insertlinkname.link)."

Jimmy's Artifact:	Artifact Description:
	This artifact is a clip from a blog in 2035, talking about the earliest AI text-to-image artistry we know of. According to this blog, the user could input a string of keywords and come up with a piece of art that matches it based on data found on google or input from the programmer.

Just as with June's fiction story above, this artifact draws me into Jimmy's imagined world, so I can share in their vision of what the future might look like. In Jimmy's blog post (which features a character named Jimmy, apparently their fictional name of choice for many purposes), they imagine an AI that can generate art based on keywords. Jimmy then discusses a teen trend on "ClikClok" where users input personal information to generate art. Interestingly, the "backlash" to this is, in this short fiction, from artists whose livelihood is threatened rather than, say, parents objecting to breaches in privacy or teens using the tool to generate harmful images with others' personal data. In Jimmy's notes and dystopian scenario, Jimmy does talk about this risk of misuse and harm via deepfake, so perhaps they were exploring a middle ground where the absolute worst that Jimmy imagined in their scenario is not happening.

Crucially, too, teens are not entirely villainous in this story; they are not abusing the tool or misusing others' data. This bucks a common trend in popular discourse around teens and technology, where people often fear the worst, mistrusting, punishing, and surveilling teens in their technology use. As an instructor, and literacy scholar, I value Jimmy's counter-text to the

prevailing discourse around young people and technology. I see this text as a way that Jimmy is exploring ethical issues with their chosen technology, but also resisting deficit-based portrayals of teens and technology. Instead of being the source of harm, the ClikClok teens are using the AI to "see what art it makes" with their information, perhaps exploring ways that data and AI can reflect or distort aspects of their identities. But from Jimmy's other writing, we know there is the looming threat of someone abusing this capability to create deepfakes. The artifact hangs in the delicate balance between exploration and creativity on the one hand and potential grave harm—to artists and to victims of deepfakery—on the other. Engaging in speculation and writing their own version of the future gave Jimmy an opportunity to think with ethical theories but also exercise agency (again, we see that agency and ethics are deeply intertwined) by crafting the narrative in a certain way.

Although I wish I could report on how completing this project might have changed the way Jimmy engaged with the discourse that subsequently exploded over AI art, I know that I benefited from seeing Jimmy's writing and thinking before the rest of the world started talking about these issues (Jimmy wrote this before DALL-E 2 was made free to the public, and before the current explosion in discourse around AI art). I can only imagine that this helped prepare Jimmy to contribute as well. Perhaps in 2035 our AI art generators might also include depictions of you and me, using private data—a chilling precaution we might learn from Jimmy's speculation.

Reflecting on Using Fictional AI: Finding Hope and Inspiration in Fictional AI, but Worrying about Instrumentalizing It

Just as I directly sought input from students on whether they thought learning the ethical theories was important, I asked for student perspectives on whether fictional AI is important. On day one of week four, I asked the 8th graders to respond to this pre-write prompt: "Do you think fictional AI is important for understanding the AIs we use or will use? Why/why not/how? Come

up with an example that supports your opinion." Their responses ranged from one definite "no" to some "yes and no," and many "yes" answers. Simon, our naysayer, says "I Dont think [fictional] AI is a good Representation. Most of the time it is nothing like what AI even is." I agree with Simon here that fictional AI has limited ability to, say, help us understand the functionality of a current facial recognition system, although I do think it would be a success of the unit if a student were able to point out the ways they differ. Fictional AI can, however, help us "theorize how it will exist and thrive in our environment" (Jonah). Hailey gave us several ways that fiction could have differing impacts, from dangerous to inspirational. A couple of students noted that fictional AI already has inspired or predicted real-world AI. Finally, Theo offers us a way to use fictional AI to build hope: "I think that fictional is good so we can hope/and or wish that that is something that can happen and help make AI better." Perhaps this is the agentic practice I had been seeking: using fictional AI to not only imagine a better world with AI, but work to make it so.

However, I still bristle against the idea of a solely instrumental view of fiction, where educators use it as a tool for a learning objective they have created. Perhaps this is the result of my experience and training as a librarian: I always saw it as such a privilege to be able to eschew judgment of what or how someone might want to read. My job was to help people find what they wanted or needed—steamy romance? We have some you'll love. Murder on murder? I'll load up your basket with the grisliest. Non-fiction on an embarrassing subject? Not a problem—check it out and your record is confidential. The patron determined their own reason for seeking out a text, and my job was to help them find it without judgment. This is (part of the reason) why public libraries remain so important. In so many other parts of their lives, adults are choosing and judging texts for kids, using them with motives of their own making rather than the child's.

Of course, choosing to use a text, including fiction film and prose, for a particular purpose is also important, effective, and often joyful—that is why I chose to use this tactic in the first place. In using this strategy, though, I think it is important to be open to how students might want to use the text in ways you did not design, resisting interpretations you had anticipated and/or expressing their opinions about the text. In foreclosing text choice (even when offering options) in the curriculum, I must also make sure students know that if what I choose to watch or read does not bring them joy, there's a whole world out there, for free at their public library, that might.

Chapter Conclusion

My theory was that multiple access points through different modes of thinking and different disciplinary perspectives might open up access to AI to students with all kinds of prior experience. For example, a science fiction buff could offer their expertise on *Star Trek*, while a computer science enthusiast could offer theirs on tweaking a program, and both would be valued in the space of the class. In this chapter, I focused on the science fiction part of my pedagogy, using examples of a film discussion, writing prompts, and the analysis of a student-produced text to show the affordances and shortcomings of teaching and learning with science fiction consumption and production.

In these examples, I saw a few major ways that fictional AI served students: first, it served as a foil for "real" AI, allowing us to ask questions about and explore the limits of current-day AI that impacts us; second, it allowed us to explore ethical implications of technology in a cheap, low-risk, and fun way; finally, it gave students an opportunity to prepare for the future. For example, the B-Bots in *Ron's Gone Wrong* can help students think about how their school governs their use of phones, and Jimmy's AI art speculation might help them consider how they treat art generated by DALL-E 2. Learning about the current technical capabilities and functioning of AI applications—like how training data affects an AI's functionality—can help us

imagine better ways of building it in our fiction and back again in the “real world.” Stories like June’s can show how students think through the relative risks and benefits of a particular AI, and how these risks and benefits might play out when the technology functions as designed and when it fails. I would have liked to incorporate more student responses to their peers’ fiction stories, if I had more time. That way, students could not only express their own ideas about risks, benefits, and future designs, but also apply some of the ethical lenses to interpreting their peers’ stories.

I also recognize my own tendencies to simultaneously embrace and resist the instrumentalization of fiction, using its production and consumption as a means to another learning objective. This does not mean that it is not useful for this purpose, just that it warrants a reminder that there are other ways that students do and should experience fiction.

Chapter 7: Discussion and Conclusions

In Chapters 4 and 5, I discussed findings related to big questions that animated the curriculum and teaching: What and where is AI? How am I connected to it and what can I do about it? How should we live with AI? In Chapter 6, I focused more closely on the function and design of teaching with science fiction. In this chapter, I will revisit the findings, then draw connections among them, discussing three challenges that cut across all of the findings sections. Next, I revisit the theorization of critical posthumanist literacy theory, discussing the implications of this study for developing pedagogy and further questions that arise from the study. Finally, I will conclude by revisiting Cynthia's story and AI-mediated search.

Findings at a Glance

In Chapter 4, I showed how many students went from seeing AI as mostly just robots at the beginning of the unit to seeing AI everywhere, in almost everything, by the end of the unit. This was a difficult overgeneralization to address, given how hard it is to locate reliable, intelligible information about where and when AI is used, and also how what counts as AI is changing and changeable. I examined activities designed to show AI as a computational/technical process and as a speculative production, highlighting the strengths of these activities on their own but noticing that they served to reinscribe the differences in these approaches rather than bridge them as I had hoped. I found that several students and a teacher missed having a historical understanding of AI in the unit.

Finally, I looked at student responses to AI through the lens of the hyperobject. Although AI as hyperobject offers an interesting lens through which to understand the way we might feel in the face of such a diffuse, huge thing, I see the concept's power as primarily explanatory rather than instructive: it gives us a framework for considering the massive scale and near-ubiquity of AI (and its data gathering apparatuses) and for understanding why people—including

many of my participants—could not offer reasonable ways to extract themselves from AI systems, but it does not show us a way forward for making change.

In Chapter 5, I showed how students wrote about AI as something they use, but rarely discussed how it uses them, or, more specifically, their data in agentive ways. In the AI & Data Audit, many students were unconcerned about the intrusions of surveillance mechanisms in their daily lives. They described little desire and also few options for changing how their data is used, other than opting out of using a tool completely. In another activity where I asked students to create rules around data use, students emphasized the importance of consent and privacy, but were willing to sacrifice some privacy if safety were at stake. Many of the 7th graders approved of data collection by companies, but drew the line at selling this data. Next, I looked at students' expressions of the (lack of) power they felt they had, as young people and as individuals. I had hoped the students felt powerful and able to effect change in the world should they deem it necessary, but they often expressed feeling the opposite.

Overall, students did not seem to see their agency as bound up with machines, but rather as attenuated by their youth. It is important to note here that my embrace of posthumanist conceptualizations of agency does not imply that I deny the existence of human agency; rather, I see our abilities and capacities to act as individual humans as bound up in dynamic sociomaterial meshes. However, the students did not express seeing their agency this way, even when they described their choices as being affected by or affecting machine processes (for the latter, I am thinking of June's assertion that with the knowledge gained from her final project on AI in advertising, she will be able to get better ads for herself).

In the second part of Chapter 5, I discuss students' responses to instruction around ethics. This included how students applied ethical lenses and thinking practices flexibly and effectively during group discussions, without getting hung up on terminology. In addition to teaching ethical thinking practices, I was interested in what students saw as important ethical

issues, and they expressed a wide range of them as related to our future with AI, but a much narrower range (focused foremost on bad human behavior) when asked about current tech issues. I relate how some students' enthusiastic and skeptical attitudes toward AI persisted, but that students remained overwhelmingly positive about AI despite giving them too few opportunities as a class to imagine what a good life with AI might look like. Finally, I discuss how peers with diverse opinions served as important thought/conversation partners when engaging in ethical thinking, pushing and pulling each other in discussion and offering up their expertise when possible.

In Chapter 6, I wrote about how I used science fiction consumption and production as a core part of my pedagogy. I described three functions that speculative fiction activities served: as a foil for thinking about our everyday reality, in the case of a discussion around science fiction film *Ron's Gone Wrong* (Smith & Vine, 2021); as an opportunity to conduct ethical thought experiments, as illustrated by the case of June's science fiction story; and as practice or preparation for the future, looking at Jimmy's prescient final project on AI in art. Although I see fiction as an effective, and often joyful, means to instructional ends, I also cautioned that only treating it as instrumental in achieving other goals can have unintended ill effects.

Synthesizing across Findings Areas

One of the most difficult things to do in the preceding chapters was to draw boundaries among these ideas and questions, deciding how to separate out the concepts and findings. The difficulty I had in deciding where each should go is an indication of their interconnectedness; where you see AI is bound up with how you understand what AI is; where you see AI, and yourself as connected to it, is linked to how you might see your own agentive capacity; and how you think of your capacity to act is deeply tied to what you think of as your duty to act. Disentangling these ideas temporarily does a couple kinds of work: for the reader, it allows the ideas to be clearer and more logically presented, and for me, the analyst, it provides guidance

for which data to include as I seek patterns and connections to theory in my analysis. However, these are logistical and analytical tools rather than a reflection of the distinctness of the concepts.

This chapter is a chance to weave together my guiding concepts, findings, and theory. It is also an occasion to explore themes that traversed the findings, like how I wrestled with teaching an unsettled, sometimes unsettling topic, how I navigated finding and troubling the boundaries between "real" and imagined AI, and how I tried to weave together different disciplinary ideas. In this section, I discuss three cross-cutting challenges:

1. Finding the balance of "just enough" for interdisciplinary instruction
2. Falling back into humanist habits
3. Teaching with uncertainty

Although in this section I center challenges, this does not mean that there were not any cross-cutting strengths. In particular, critical literacy practices proved especially well-suited to incorporate into designs where we (perhaps obviously) critiqued AI, asking questions about how it is constructed/authored, and built our own "counter-texts" through speculative fiction. I will discuss expanding these applications of critical literacies practice in the section detailing implications for pedagogy.

Challenge 1: Finding What Is "Just Enough" for Interdisciplinary Instruction

This curriculum, and this scholarship more generally, spans different disciplinary traditions, including literacy, computer science, and ethics/philosophy. There are endless permutations of how to combine these, drawing on their respective theories, practices, and histories. My job in designing and implementing this curriculum was to develop *a* way, try it out, evaluate it, and redesign it. This meant that at every step, I was thinking about what and how to include from each of these three disciplines, two of which are not my home territory. When thinking about what and where AI is in the world, and in these students' lives, I used technical

activities developed by computer scientists (e.g., Payne, 2019; Webster, 2019), speculative fiction activities, and examples of ethical dilemmas, along with activities designed to see across the perspectives (e.g., collecting definitions and examples of AI on Jamboard). However, I was never sure if, by incorporating these multiple perspectives, each one got enough attention; was I sacrificing breadth for depth? My attempt to answer this question involved trying to determine what might be "just enough" exposure to disciplinary knowledge and traditions at the right time. Taking the computational perspective first, I had to make choices given constraints on instructional time and my own expertise. The Teachable Machine activity is a good example of this. I wanted to use Teachable Machine to demonstrate important machine learning concepts, show AI as a computational process (AI ontology), allow students an opportunity to create and tinker with AI firsthand (entangled agency), and connect technical knowledge, particularly relating to the effects of bad or biased training data, to broader ethical issues (ethics & justice). I think this activity served those purposes, and was "just enough" to accomplish these goals, but at the expense of deeper engagement with the underlying code.

Students and educators are also interested in finding the boundaries of "real" AI, separating out what is currently possible from what is not, and which technologies actually use AI from those that do not. The latter is even listed as a core competency of AI literacy as described by Long & Magerko (2020). There is a certain amount of technical exposure students need to be able to make these determinations. I liken the exploration of how much technical expertise is necessary to be critical about AI to what you might need to know about film production; understanding how a film is made, with actors, a script, and hardware and software is crucial to media literacy practices like understanding that media messages are constructed (Kellner & Share, 2005). This does not mean that students must know every detail of how the camera works, but being able to separate out "reality" from "fiction" in these contexts requires

some knowledge of the technical arrangements—editing software, digital recorders—and the social ones—actors, scripts, directors—involved in producing and distributing a film.

We are still learning the limits and possibilities of AI, and there are many barriers—transparency, trade secrets, unexplainability—to developing a "just enough" pedagogy that will allow students to critically make "cuts" about what is and is not AI. However, this does not absolve us (educators) of the need to try. Ultimately, it is important to note that interdisciplinary curriculum is not a replacement for disciplinary exposure, but rather should be a supplement to it. "Just enough" technical knowledge and experience for doing critical posthumanist literacy work is just that; it is not enough to, say, train an engineer. It is the meaningful minimum to serve a particular purpose, working toward sociotechnical justice through interdisciplinary critical posthumanist literacy practices.

Challenge 2: Falling Back into Humanist Habits

In Chapter 2, I describe my framework as reaching toward posthumanism from a base of critical literacy theory and praxis. In my analysis, I realized that I often fell short of fully integrating posthumanist concepts into my teaching and design. Looking at my instruction around ethics, for example, I see missed opportunities to explicitly ask the students to consider the ethical assemblages that we and AI form. Of course, this also connects to AI ontology by drawing on what makes up AI, as assemblage, in the first place. Incorporating more posthumanism in teaching around ethics could look like something as simple as adding a posthumanist lens to the framework with other (dominant) ethics lenses, asking students to consider dilemmas from this standpoint as well. We could use Bennett's (2010) ideas about choosing ethical assemblages that tend towards enacting good rather than harm as an option for a posthumanist ethical lens. Perhaps we could ask students to think with this idea of the ethical assemblage, first mapping out what it might look like or contain, then evaluating whether they think it might bend toward doing good or doing harm.

This study is not unique in the failure to be fully posthumanist; it is a long-held critique of posthumanism that we can never truly escape the limits of our human perceptions, our human thinking machinery, and our tendencies to center ourselves even when we set out not to. One example of the constraint of our humanness on this work could be (human) language itself. Do we, and our students, have enough language to express interrelationships in actor-networks? Most students said they were connected to AI because they "use it." Does the subject-object construction common in our everyday language and present here interfere in some way with our ability to imagine a different, more complicated relationship?

Finally, I wanted, from the outset, for students to feel agentive, or to "open up agentive possibilities" through the curriculum. This is perilously close to the humanist impulse to "empower" students, long a goal of certain educational programs, but problematic in its conception of power as something to give and possess (among other challenges). However, I found myself still unable to reconcile attenuated notions of agency from posthumanism (as, for example, distributed and emergent, rather than possessed by an individual) with the goal of opening up possibilities for meaningful participation and effecting change for each individual student. I also found myself linking agency to hope, thinking that if students did not feel agentive they would also feel hopeless about the future with AI. Although I think this was sometimes the case, it was definitely not true all the time. Despite seeing major barriers to effecting change, students were overwhelmingly hopeful about AI's promise for the future. The way I see through this challenge, at least temporarily, is to think of hope as relational rather than rational. Rather than tethering hope to some idea of how much power we feel, we can think of it as something that lives between the people who surround us. If a student assesses their ability to make systemic change as negligible (and perhaps it is), we can shift the focus to the ways they can feel in control and doing good for each other and the world. Pinning all our hope on systemic change is too risky a bet.

Challenge 3: Teaching with Uncertainty

AI can be a slippery subject; it is shape-shifting and invisible, simultaneously powerful and laughably dysfunctional, seemingly everywhere and nowhere at once. This can make it hard to investigate with students and design around as a teacher. Any activity designs must be general enough to adapt to changes in technology and the needs and interests of the students, but still specific enough to give purchase for criticality. Although I put a lot of stock in design and redesign of materials and activities, there are distinct limits to any ready-made curriculum. The magic happens at the nexus of design and implementation, in the messy fray of the classroom or other learning environment.

One specific area of uncertainty that appeared repeatedly in the findings was how, when, and why we might draw sharp distinctions between "real" and imagined AI, as I also discussed above in connection to finding the ideal blend of disciplinary approaches. One possible way to redesign around AI ontology is to make clearer divisions between "real"/currently implemented and fictional/speculative AI, while maintaining the connections between them. This is different from my initial aim, which focused more on blurring the boundaries rather than reinscribing them. In practice, identifying what was currently possible, versus fanciful or scare-mongering, became a sort of de facto marker of expertise (linking the practice of defining AI to a student's agency as relatively expert, or as able to make distinctions). I felt a heavy responsibility with teacherly authority (but little computational expertise) to determine what counted as "real AI." Many teachers will likely be in this same position, so it is worth examining. Is the ideal here having more computer science training for all teachers? (But what if the training becomes rapidly out-of-date with the advent of new, specific technologies and their uses?) Or perhaps we give up on firm answers and focus more on the process of researching these answers? (But the answers change!) Knowing what the

boundaries of “real” AI is difficult, if not impossible, but also very useful in critical pedagogy; it is helpful to be able to say whether or not, for example, your phone is listening to you and also know how to find reliable information to answer that question yourself. However, this is not to diminish the interconnection of imagined and “real” AI, as there are obvious benefits to speculating about the technical capacities and applications of current and future AIs. But it is powerful to be able to determine when, say, a company is bluffing about what their AI can or does do (for example, hiding the role of humans in the process) or about the data privacy practices they employ, just as it is powerful to imagine the next AI to combat climate change.

Implications for Pedagogy

In Chapter 2, I explored my theoretical framework, critical posthumanist literacy, and showed how it draws from critical literacies and posthumanist theory and practice. Here, I think with these ideas again to evaluate their strengths and weaknesses for teaching and learning about AI, returning to the most fundamental question of the project: What and how should we teach about/with AI in order to improve human relationships with each other, the Earth, and nonhumans?

Looking across the findings areas, I can identify several important aspects of a critical posthumanist literacy pedagogy. The pedagogy must, in addition to the other ideological aims of critical literacies pedagogy (like democratic participation, social justice), account for uncertainty and change and embrace practices that generate and test speculation. It was also tremendously important to include not only different disciplinary perspectives, but also to draw on the diversity of student interests and perspectives. Allowing students to choose topics and formats for completing assignments was an easy, and impactful, way to honor those interests (just as allowing students to choose the texts they would like to read does the same). Though these may seem obvious, it is important that critical posthumanist literacy curricula are flexible, responsive to student interests and needs, along with changing technology; that they

incorporate opportunities for hopeful and cautionary speculation; that they allow for student choice where possible, as a (bare minimum) way to honor student agency in a curriculum that might expose limits to it.

Furthermore, in the absence of (felt or practiced) student agency, we have tremendous responsibility for making sure we get data privacy protection in schools right and also that we build opportunities for students to exercise choice and practice criticality in our instructional practice. Although it is not going to change a company's data privacy practices, allowing students to choose a project topic, which they then feel expert enough on to shape the ads that they see (using the example of June), is consequential. Data privacy education, although not always a perfect or easy fit for AI ethics curriculum, belongs here—we just need to work out how best to fit it meaningfully in. It cannot remain exclusively in the grip of punitive curricula that focus on student safety (or, more probably, danger), giving them long lists of "don'ts" at the expense of broadening our views of student agency and participation in complex sociotechnical systems.

Critical Posthumanist Literacy, from Propositions to Practices

In Chapter 2, I set out a framework for critical posthumanist literacy, listing five "propositions" that undergird it, summarized again here:

1. AI is a part of posthumanist literacy practice.
2. We can use ideas developed in critical literacies to analyze/improve everyday AIs, although no critical literacy model on its own is sufficient to do so.
3. We can use concepts from posthumanist theorists to extend critical literacy practice into the age of machine writing/reading/living.
4. Critical literacy and posthumanist theories, even when combined, still must be combined with (at least) "just enough" computational/technical know-how to be most useful.

5. Critical literacies have a rich tradition of pedagogical theory and practice that can be helpful in enacting posthumanist ideas.

To do critical posthumanist literacy, we must recognize AI's role in our literacy practices, use strengths of critical literacy praxis and posthumanist thinking, and incorporate enough technical experience for students to be able do the kinds of critical, participatory, and creative work that might lead to more ethical assemblages/sociotechnical justice vis-à-vis AI. In practice, this means that we would teach students basic technical principles of AI that would allow them to make meaningful critiques and decisions about their uses of/by AI.

I offer an initial list of ideas for core practices that we might consider as part of critical posthumanist literacy in the box below. One of these practices, for example, is to "ask questions about how an AI functions and connect these to social impact." In this study, we engaged in this practice when we trained models in Teachable Machine during the unit, learning that if we spend more time training on a particular class of data, or have something in the background that shouldn't be there, our model will, predictably, malfunction. Then we imagined the impact of these kinds of biases or errors in training if we were not making a banana ripeness detector, but rather an AI that must identify a face. Educators can also help students see when AI processes are guiding what they see, on YouTube, say, and encourage them to investigate how the recommendation process works. Even if students and teachers cannot answer this question definitively (as is often the case with proprietary AIs), we are cultivating human agentic practices simply by investigating, figuring out what the limits of our understanding are.

Key Critical Posthumanist Literacy Practices:

1. Identify when AI is operative in your daily lives, paying special attention to your communication practices—including reading/consuming media, writing/composing, information seeking—and how it operates in your community, and beyond
2. Ask questions about how an AI functions and connect these to social impact
3. Identify—and join, where possible—ethical assemblages that tend toward sociotechnical justice

4. Explore your (human) agency as constrained and enabled by AI/data collection and surveillance
5. Think of AI critically as both a text (non-neutral, constructed) and as an actor/agent (agentive, shaping behavior)
6. Explore participation with AIs—create, test, critique, challenge them

These practices, of course, are not designs. They are suggestions for working with students toward engaging with AI as a consequential part of literacy practice. Furthermore, translating them into curriculum and activities is not a straightforward operation: just as critical literacy scholars caution against prescriptive curricula, critical posthumanist literacy scholars should too. How these practices can and should be enacted will vary widely depending on the needs and interests of the particular humans who are teaching and learning.

Putting Critical Posthumanist Literacy Practices into Schools

While one goal of this study was figuring out how to broaden access to AI, from a critical and literacy-based perspective, it cannot be replicated at scale. This unit, in its current form, was designed to take many hours of instructional time, without any standardized testing gains to show for it. I taught in a private school because of COVID-19 restrictions in our public schools, but I also imagine it would have been difficult to find a space for the unit in a public school schedule (because of time constraints and disciplinary separation of content/curricula). This leaves us with a big challenge: how can more students have critical access to learning about/with AI? Where might this kind of instruction fit in the standard (public) school day? How can we fit in with/skirt current systems of accountability-via-testing that drive public schooling in America?

Perhaps we can use the current fervor around and interest in AI (and even AI in education, now) to build critical literacy practices into our computer science curricula (already happening in some places, like New York (Singer, 2023) and to bring “just enough” technical information into our critical literacy spaces to have informed conversations about AI there, too. In the absence of time and will to devote entire weeks of instruction to interdisciplinary

curriculum around AI and ethics, we can work on ways that computer science and literacy can strengthen each other, in both class periods. This means that, hopefully, more students will have access to learning about AI from different perspectives, and also will have the opportunity to explore critical ideas with a really interesting technological system. In other words, we might find that work at the intersection strengthens both. Perhaps, then, it is a blessing in disguise that no activity I did truly blended perspectives; in our disciplinarized education (and perhaps work) world, we might still need to rely on the paradigm of porting other disciplines' practices in rather than fusing them together fundamentally.

The impulse to bring AI into critical literacy pedagogy and scholarship is borne of what I see as necessity; AI is already a part of our literacy practices. However, it also has implications for how public education can take up AI education.

Further Questions

This was a first foray in using these techniques for critical posthumanist literacy pedagogy. There is significant work to be done, and, at this point, more lingering questions than answers. One of these questions is how we "open up agentive participation" while thinking with Barad's intra-active agency (2007) and Bennett's distributed agency (2010). These give us ways to understand our own agency as bound up in bigger systems or collections of things. In my estimation, this mostly has the effect of forcing us to revise downward our estimates of our own (individual) capacity to act, as we think of how this capacity is distributed (Bennett's term) among humans and nonhumans. I think students already feel this diminution of their sense of agency, if not because they are considering the enormity of AI-data systems, but because they see their youth as a hindrance to powerful participation in the world. I fear what I did was give them access to this other way of shrinking their (felt) agency, where I showed them, for example, how what they see in their feeds (in social media, YouTube, Netflix) is chosen for them by AI processes, how resume-reading AIs might prevent them from getting a job, among a

host of other ways AI constrains what we see, read, and perhaps even think. This makes me wonder, is AI fundamentally a (human) agency-constraining object? How might we build it, or more urgently, teach about it, in ways that open us up to other perspectives (not serve us pre-selected ones meant to anger us for more "engagement") and give us good options for seeking justice? There is a fundamental tension between the impulse to make students feel powerful and the application of posthumanist theories that decenter and diminish this power (however rightly).

Awareness and analysis are only the beginning of critical digital literacy, which emphasizes the role of active participation in technical systems. So too are the goals of critical posthumanist literacy ultimately about student participation in and with AI systems in (human) agentive ways. However, there are still several things to think through in designing for agentive AI participation: how much awareness is necessary for critical participation, and must teaching in this vein precede it? What does agentive participation look like for children vis-à-vis large AI models? I suggest that we start by employing cycles of tinkering and critiquing; when we use and/or build AI models (with Teachable Machine, for example), what breaks and why? What might the ethical implications of both the functioning and failures of this model be? We began this kind of inquiry when we explored Teachable Machine, and I demonstrated the banana ripeness detector. I asked the students to imagine if we weren't detecting bananas, but humans—what would it mean for the AI to function improperly then? However, this was an exercise—a fun one, to be sure—in a sort of proof of concept (about how training data can shape the functionality of your AI, and it is easy to mis-train it so that it gives you incorrect or biased results). What would it mean for students to develop real-world, critical participation with AI systems beyond "toy" examples in the classroom?

Given the nascent stage of empirical research around AI and ethics curricula, I also wonder, How might this kind of learning change students, and technology? I do not mean this in

a direct, or quantifiable, cause-and-effect kind of way, as people and technology are far too complicated, and agency too diffuse, to be reduced to these kinds of simplistic narratives. But even without a direct, reductive line, we can think about ways that students' beliefs and actions might shift as a result of accumulating exposures to critical posthumanist literacy practices, and perhaps even affect the way that they use—or perhaps develop—technology themselves. My findings here indicate that writing and viewing speculative fiction is a joyful, and valuable, way to explore issues in AI ethics with middle school students, but it does not necessarily deepen their technical knowledge. However, I wonder about the effects of speculative pedagogy on technical designs; if software engineers engaged in speculative fiction writing, and especially more ethical experimentation in this format, would or could this affect the design of technology?

The final, and impossible, question is a speculative one: What will the future hold, and how might we shape it? The answer coursing through this dissertation is that teaching and learning, joyfully, seriously, respectfully with children is key.

Conclusion: Coming Full Circle to Search

As I have been writing this dissertation, stories about ChatGPT and internet searching have started appearing in the news. As I was writing, on February 13, 2023, *Nature* published a "News Explainer" on their website titled "AI chatbots are coming to search engines – can you trust the results?" (Stokel-Walker, 2023). The article reviews news about the move toward using chatbots to deliver search results rather than seeing a list of sites. The results are still unreliable, even with the most straightforward questions, but money is flowing toward the implementation of these AI searchbots. This brief article points to ethical issues beyond reliability, like lack of transparency (about how the language model is trained, for example) and misplaced (human) trust in results.

By asking students how we might build a better future with AI, I hope to highlight how none of this is inevitable, if we only start asking the right questions and using the answers we

find to change our actions. In the case of search, we should ask, What happens to human agency when AI gives us a singular answer (and we treat it as the only/right one)? Who gets to design, train, and maintain the tool that supplies our single answers? Who teaches the students how to critique these answers or build a new way to find information, should they need it? I cannot help but wonder, what would this kind of search have done for Dylann Roof? Could a difference in the way AI shaped his reading have somehow altered his decisions? And what if he had been taught to be critical about the way that information was produced and shared? These are haunting questions, but we do not need to (nor can we possibly) know the answers to them to understand the urgency of reaching and teaching toward sociotechnical justice, impelling students to recognize and ask questions about how technology shapes what we read, write, and think—and vice versa.

References

AI effect. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=AI_effect&oldid=991136011

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=pc1OkzviktiLco0hJY5BDRPpI44H_bKV

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Behdadi, D., & Munthe, C. (2020). A normative approach to artificial moral agency. *Minds and Machines*, *30*(2), 195–218. https://doi.org/10.1007/s11023-020-09525-8

Behrman, E. H. (2006). Teaching about language, power, and text: A review of classroom practices that support critical literacy. *Journal of Adolescent & Adult Literacy*, *49*(6), 490–498. https://doi.org/10.1598/JAAL.49.6.4

Bennett, J. (2010). *Vibrant matter: A political ecology of things*. Duke University Press.

Blackford, R. (2017). *Science fiction and the moral imagination :Visions, minds, ethics*. Springer International Publishing.

Boaz, C. (2020). How speculative fiction can teach about gender and power in international politics: A pedagogical overview. *International Studies Perspectives*, *21*(3), 240–257. https://doi.org/10.1093/isp/ekz020

Bostrom, N., & Yudkowsky, E. (2011). The ethics of artificial intelligence. In W. Ramsey & K. Frankish (Eds.), *Cambridge Handbook of Artificial Intelligence*. Cambridge University Press. https://doi.org/10.1016/j.mpmed.2010.10.008

Boyd Waters, M., & Myers. (2023, February 27). ChatGPT in English class: Using AI writers as thinking partners. *ReThink ELA*. https://www.rethinkela.com/2023/02/chatgpt-in-english-class-using-ai-writers-as-thinking-partners/

Braidotti, R. (2019). A theoretical framework for the critical posthumanities. *Theory, Culture & Society*, *36*(6), 31–61. https://doi.org/10.1177/0263276418771486

Bratton, B. H. (2015). *The stack: On software and sovereignty*. MIT Press.

Buchholz, B. A., DeHart, J., & Moorman, G. (2020). Digital citizenship during a global pandemic: Moving beyond digital literacy. *Journal of Adolescent & Adult Literacy*, *64*(1), 11–17. https://doi.org/10.1002/jaal.1076

Byrne, J., Day, E., & Raftree, L. (2021). The case for better governance of children's data: A manifesto. UNICEF Office of Global Insight and Policy. https://www.unicef.org/globalinsight/media/1741/file/UNICEF%20Global%20Insight%20Data%20Governance%20Manifesto.pdf

Carrington, V. (2018). The changing landscape of literacies: Big data and algorithms. *Digital Culture & Education*, *10*(1), 67–76.

Chao, J., Finzer, B., Rosé, C. P., Jiang, S., Yoder, M., Fiacco, J., Murray, C., Tatar, C., & Wiedemann, K. (2022). StoryQ: A web-based machine learning and text mining tool for K-12 students. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2*, 1178. https://doi.org/10.1145/3478432.3499267

Cobb, P., Lehrer, R., Schauble, L., Confrey, J., & DiSessa, A. (2007). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13. https://doi.org/10.3102/0013189x032001009

Coeckelbergh, M. (2020). *AI ethics*. MIT Press.

College Board. (2020). *AP computer science principles course and exam description*.

Collin, R. (2020). Four models of literature and ethics. *English Journal*, 9.

Common Sense Media. (2023a). *Common Sense privacy evaluations*. https://privacy.commonsense.org/evaluations/1

Common Sense Media. (2023b). *Digital citizenship*. https://www.commonsense.org/education/digital-citizenship

Dalton, B. (2021). *30-second stories*.

Dixon-Román, E., Nichols, T. P., & Nyame-Mensah, A. (2020). The racializing forces of/in AI educational technologies. *Learning, Media and Technology*, *45*(3), 236–250. https://doi.org/10.1080/17439884.2020.1667825

Ellis, R., Rose, C., Chao, J., Jiang, S., Finzer, W., Wiedemann, K., Tatar, C., & Fiacco, J. (2022). *AI in ELA: A DEI solution*.

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St Martin's Press.

Feathers, T. (2019, August 20). *Flawed algorithms are grading millions of students' essays*. Motherboard: Tech by Vice. https://www.vice.com/en/article/pa7dj9/flawed-algorithms-are-grading-millions-of-students-essays

Fieser, J. (n.d.). Ethics. In *Internet Encyclopedia of Philosophy*. Retrieved April 29, 2021, from https://iep.utm.edu/ethics/

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2020). *Preparing for life in a digital world: IEA International Computer and Information Literacy Study 2018 international report*. Springer International Publishing. https://doi.org/10.1007/978-3-030-38781-5

Franklin, J., & Olson, E. (2022, November 28). The Buffalo Tops shooting suspect pleads guilty to state murder charges. *NPR*. https://www.npr.org/2022/11/28/1138700312/buffalo-tops-shooter-guilty-plea-state-charges

Freire, P. (1970). *Pedagogy of the oppressed* (30th anniversary ed). Continuum.

Garcia, A., Mirra, N., Morrell, E., Martinez, A., & Scorza, D. (2015). The council of youth research: Critical literacy and civic agency in the digital age. *Reading & Writing Quarterly*, *31*(2), 151–167. https://doi.org/10.1080/10573569.2014.962203

Gleick, J. (2011). *The information: A history, a theory, a flood*. Vintage Books.

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.

Harman, G. (2018). *Object-oriented ontology: A new theory of everything*. Pelican.

Heilweil, R. (2019, December 12). *Artificial intelligence will help determine if you get your next job*. Vox. https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen

Hickey-Moody, A. (2009). Little war machines: Posthuman pedagogy and its media. *Journal of Literary & Cultural Disability Studies*, *3*(3), 273–280. https://doi.org/10.1353/jlc.0.0024

Hinchliffe, G. (2000). Education or pedagogy? *Journal of Philosophy of Education*, *35*(1), 31–45. https://doi.org/10.1111/1467-9752.00208

Holland, D., Skinner, D., Lachicotte Jr., W., & Cain, C. (1998). *Identity and agency in cultural worlds*. Harvard University Press.

Hollin, G., Forsyth, I., Giraud, E., & Potts, T. (2017). (Dis)entangling Barad: Materialisms and ethics. *Social Studies of Science*, *47*(6), 918–941. https://doi.org/10.1177/0306312717728344

Hull, G. A., & Stornaiuolo, A. (2014). Cosmopolitan literacies, social networks, and "proper distance": Striving to understand in a global world. *Curriculum Inquiry*, *44*(1), 15–44. https://doi.org/10.1111/curi.12035

Hundley, M. (2021). *QCQC*.

Hunter, T. (2021, November 14). Ask help desk: No, your phone isn't listening to your conversations. Seriously. *Washington Post*. https://www.washingtonpost.com/technology/2021/11/12/phone-audio-targeting-privacy/

IEEE. (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*. https://standards.ieee.org/content/dam/ieee-

standards/standards/web/documents/other/ead1e.pdf?utm_medium=PR&utm_source=Web&utm_campaign=EAD1e&utm_content=geias&utm_term=undefined

Is TikTok a harmless app or a threat to U.S. security? (No. 11). (2020, November 15). In *60 Minutes*. https://www.cbsnews.com/news/tiktok-cybersecurity-china-60-minutes-2020-11-15/

Jackson, A. Y., & Mazzei, L. A. (2012). *Thinking with theory in qualitative research: Viewing data across multiple perspectives*. Routledge.

Jackson, A. Y., & Mazzei, L. A. (2013). Plugging one text into another: Thinking with theory in qualitative research. *Qualitative Inquiry*, *19*(4), 261–271. https://doi.org/10.1177/1077800412471510

Johnstone, J. (2007). Technology as empowerment: A capability approach to computer ethics. *Ethics and Information Technology*, *9*(1), 73–87. https://doi.org/10.1007/s10676-006-9127-x

Jones, R. (2019). The text is reading you: Teaching language in the age of the algorithm. *Linguistics and Education*, 100750. https://doi.org/10.1016/j.linged.2019.100750

Kellner, D., & Share, J. (2005). Toward critical media literacy: Core concepts, debates, organizations, and policy. *Discourse*, *26*(3), 369–386. https://doi.org/10.1080/01596300500200169

Kuby, C. R., Spector, K., & Thiel, J. J. (2019). Cuts too small: An introduction. In C. R. Kuby, K. Spector, & J. J. Thiel (Eds.), *Posthumanism and literacy education: Knowing/becoming/doing literacies*. Routledge.

Lambert, P. (2018, May 8). *Write emails faster with Smart Compose in Gmail*. Google. https://blog.google/products/gmail/subject-write-emails-faster-smart-compose-gmail/

Lanier, J., & Weyl, G. (2020, March 15). AI is an ideology, not a technology. *Wired*. https://www.wired.com/story/opinion-ai-is-an-ideology-not-a-technology/

Larsen–Freeman, D. (2019). On language learner agency: A complex dynamic systems theory perspective. *The Modern Language Journal*, *103*(S1), 61–79. https://doi.org/10.1111/modl.12536

Latour, B. (2005). *Reassembling the social: An introduction to Actor-Network-Theory*. Oxford University Press.

Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems Practice*. https://doi.org/10.1007/BF01059830

Leander, K. M., & Burriss, S. K. (2020). Critical literacy for a posthuman world: When people read, and become, with machines. *British Journal of Educational Technology*, *51*(4), 1262–1276. https://doi.org/10.1111/bjet.12924

Lewison, M., Flint, A. S., & Van Sluys, K. (2002). Taking on critical literacy: The journey of newcomers and novices. *Language Arts*, *79*(5), 12.

Lin, P. (Director). (2015, December 8). *The ethical dilemma of self-driving cars*. TED-Ed. https://www.youtube.com/watch?v=ixIoDYVfKA0

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

Luke, A. (2012). Critical literacy: Foundational notes. *Theory Into Practice*, *51*(1), 4–11. https://doi.org/10.1080/00405841.2012.636324

Luke, A. (2018). Digital ethics now. *Language and Literacy*, *20*(3), 185–198. https://doi.org/10.20360/langandlit29416

Matarese, J. (n.d.). *Is your phone listening, and targeting you with ads?* Https://Www.Wbrc.Com. Retrieved February 27, 2023, from https://www.wbrc.com/2022/05/27/is-your-phone-listening-targeting-you-with-ads/

McGinty, J. (2020, May 18). Project to introduce middle schoolers to AI. *UGA Today*. https://news.uga.edu/project-introduces-middle-schoolers-to-ai/

Mercer, S. (2011). Understanding learner agency as a complex dynamic system. *System*, *39*(4), 427–436. https://doi.org/10.1016/j.system.2011.08.001

Miller, E. R. (2016). The ideology of learner agency and the neoliberal self. *International Journal of Applied Linguistics*, *26*(3), 348–365. https://doi.org/10.1111/ijal.12129

Miller, J. (2020, September 18). Is an algorithm less racist than a loan officer? *The New York Times*. https://www.nytimes.com/2020/09/18/business/digital-mortgages.html

Mirra, N., Morrell, E., & Filipiak, D. (2018). From digital consumption to digital invention: Toward a new critical theory and practice of multiliteracies. *Theory into Practice*, *57*(1), 12–19. https://doi.org/10.1080/00405841.2017.1390336

MIT Media Lab Personal Robots Group, & MIT STEP Lab. (2021). *DAILy workshop*. https://aieducation.mit.edu/daily/

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 2053951716679679. https://doi.org/10.1177/2053951716679679

Morton, T. (2013). *Hyperobjects: Philosophy and ecology after the end of the world*. University of Minnesota Press.

Murphy, P. (1996). Defining pedagogy. In P. F. Murphy & C. V. Gipps (Eds.), *Equity in the classroom: Towards effective pedagogy for girls and boys*. UNESCO. https://us.corwin.com/sites/default/files/upm-binaries/32079_Murphy(OU_Reader_2)_Rev_Final_Proof.pdf

Nichols, T. P., & LeBlanc, R. J. (2020). Beyond apps: Digital literacies in a platform society. *The Reading Teacher*, *74*(1), 103–109. https://doi.org/10.1002/trtr.1926

Nichols, T. P., & LeBlanc, R. J. (2021). Media education and the limits of “literacy”: Ecological orientations to performative platforms. *Curriculum Inquiry*, 1–22. https://doi.org/10.1080/03626784.2020.1865104

Nichols, T. P., LeBlanc, R. J., & Slomp, D. (2021). Writing machines: Formative assessment in the age of big data. *Journal of Adolescent & Adult Literacy*, *64*(6), 712–719. https://doi.org/10.1002/jaal.1160

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Ochs, E. (1979). Transcription as theory. In *Developmental Pragmatics* (pp. 43–72). Academic Press.

OpenAI. (2023). *Introducing ChatGPT*. https://openai.com/blog/chatgpt

Pandya, J., & Ávila, J. (2013). *Moving critical literacies forward: A new look at praxis across contexts*. Routledge.

Pangrazio, L. (2013). Young people and Facebook: What are the challenges to adopting a critical engagement? *Digital Culture & Education*, *4*(3), 34–47.

Pangrazio, L., & Selwyn, N. (2019). ‘Personal data literacies’: A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, *21*(2), 419–437. https://doi.org/10.1177/1461444818799523

Payne, B. (2019). *MIT AI ethics education curriculum*. Google Docs. https://docs.google.com/document/d/1e9wx9oBg7CR0s5O7YnYHVmX7H7pnITfoDxNdrSGkp60/edit?usp=embed_facebook

Perrotta, R. (2023). Peabody EDI Office responds to MSU shooting with email written using ChatGPT. *The Vanderbilt Hustler*. https://vanderbilthustler.com/2023/02/17/peabody-edi-office-responds-to-msu-shooting-with-email-written-using-chatgpt/

Rautio, P., Spector, K., & Thiel, J. J. (2019). Theory that cats have about swift louseflies: A distractive response. In C. R. Kuby, K. Spector, & J. J. Thiel (Eds.), *Posthumanism and literacy education: Knowing/becoming/doing literacies*. Routledge.

Respondus (Director). (2013, October 1). *Introduction to Respondus LockDown Browser for students*. https://www.youtube.com/watch?v=XuX8WoeAycs&feature=emb_logo

Robinson, B. (2020). The ClassDojo app: Training in the art of dividuation. *International Journal of Qualitative Studies in Education*, 1–15. https://doi.org/10.1080/09518398.2020.1771460

Sabey, D. B., & Leander, K. M. (2020). More connected and more divided than ever: Toward a cosmopolitan ethics of digital literacies. In E. B. Moje, P. P. Afflerbach, P. Enciso, & N. K. Lesaux (Eds.), *Handbook of Reading Research: Vol. V*. Routledge.

Saldana, J. (2009). *The Coding manual for qualitative researchers*. https://doi.org/10.1109/test.2002.1041893

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences*, *23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Seabrook, J. (2019). The next word: Where will predictive text take us? In *The New Yorker*. https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker

Shor, I. (1999). What is critical literacy? *Journal of Pedagogy, Pluralism, and Practice*, *1*(4). https://doi.org/10.1007/978-94-6300-106-9_2

Singer. (2023). At this school, computer science class now includes critiquing chatbots. *The New York Times*. https://www.nytimes.com/2023/02/06/technology/chatgpt-schools-teachers-ai-ethics.html

Singer, P. (2011). *Practical ethics* (3rd ed.). Cambridge University Press.

Smith, S., & Vine, J.-P. (Directors). (2021). *Ron's gone wrong*. Locksmith Animation.

Sternberg, B. J., Kaplan, K. A., & Borck, J. E. (2007). Enhancing adolescent literacy achievement through integration of technology in the classroom. *Reading Research Quarterly*, *42*(3), 416–420. https://doi.org/10.1598/rrq.42.3.6

Stokel-Walker, C. (2023). AI chatbots are coming to search engines – can you trust the results? *Nature*. https://doi.org/10.1038/d41586-023-00423-4

Townson, S. (2020, November 6). AI can make bank loans more fair. *Harvard Business Review*. https://hbr.org/2020/11/ai-can-make-bank-loans-more-fair

Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Technology Law Journal*, *13*(1), 8–23. https://doi.org/10.3868/s050-004-015-0003-8

UNESCO. (2022). *K-12 AI curricula: A mapping of government-endorsed AI curricula*. https://unesdoc.unesco.org/ark:/48223/pf0000380602

Valtonen, T., Tedre, M., Mäkitalo, Ka., & Vartiainen, H. (2019). Media literacy education in the age of machine learning. *Journal of Media Literacy Education*, *11*(2). https://doi.org/10.23860/JMLE-2019-11-2-2

Velasquez, M., Moberg, D., Meyer, M. J., Shanks, T., McLean, M. R., DeCosse, D., Andre, C., Hanson, K. O., Raicu, I., & Kwan, J. (2021). *A framework for ethical decision making*. https://www.scu.edu/ethics/ethics-resources/a-framework-for-ethical-decision-making/

Webster, B. (2019, November 7). Teachable machine tutorial: Bananameter. *Medium*. https://medium.com/@warronbebster/teachable-machine-tutorial-bananameter-4bfffa765866

Webster, W. (2021). *AI-driven CCTV upgrades are coming to the "world's most watched" streets – will they make Britain safer?* The Conversation. http://theconversation.com/ai-

driven-cctv-upgrades-are-coming-to-the-worlds-most-watched-streets-will-they-make-britain-safer-157789

Wiesel, E. (1986, December 10). *Nobel Prize acceptance speech*. The Nobel Prize. https://www.nobelprize.org/prizes/peace/1986/wiesel/acceptance-speech/

Williamson, B. (2016). Digital education governance: Data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Educational Policy*. https://www.tandfonline.com/doi/full/10.1080/02680939.2015.1035758?src=recsys

Yousef, J. (2022, May 20). The Buffalo shooting suspect's online footprint prompts questions about red flags. *NPR*. https://www.npr.org/2023/02/15/1156996712/buffalo-tops-shooting-new-york-sentencing

Appendix A: 7th Grade Curriculum Overview

AI & Ethics Curriculum Overview
7th Grade

Sessions	Objectives/Topic Overview	Activity Plans
Week 1 March 23	Topic(s): ● What is AI? Objectives: ● explore AI from a variety of perspectives, identifying many examples in our world ● develop working definitions of AI for the unit ● consider connections between fictional and currently implemented AIs	● Introduction to Sarah and research project ● Overview of unit ● Assent process ● AI perspective scavenger hunt on Jamboard—find examples of AIs from different perspectives ● Writing prompt (exit ticket): What is something that you liked learning about today, something you still have a question about, or something that you want to talk about next time?
Week 2 March 30	Topic(s): ● Where is AI? ● How does AI work? Objectives: ● Explore AI from a computational perspective, training an AI model to (begin to) understand how it works, what breaks it ● Identify/locate everyday AI, surveillance & data-gathering mechanisms	● 3 parts of AI (MIT slides)* ● AI Bingo game (MIT) ● Teachable Machine activity ● UNICEF video: https://youtu.be/MBEPjFTuTPw
Week 3 April 6	Topic(s): ● Where is your data? ● How does AI classify you and others? Objectives: ● Complete personal data and AI use audits, identifying where and when AI is using our data/we use AI in our daily lives ● Develop ideas about how (students think) data should and should not be	● Writing prompt and discussion: What rules should companies, parents, and you follow when using your data? Come up with at least 3. Try to think of exceptions to your rules, too. Discuss at table groups ● Review concepts from last week (how we define AI) ● Watch and discuss (whole-group) WSJ video on TikTok (first 5 mins) ● Complete one row of audit handout together and review instructions. Give time to work on audit in table groups

	used/collected	• Discuss Google Ad settings, explore own Google Ad settings • Watch Joy Buolamwini's Gender Shades video
Week 4 April 13	Topic(s): • What are ethics and ethical lenses? • How can we use ethical lenses to think about AI? • What are the connections between "real" and fictional/imagined AIs? Objectives: • Define ethics and understand several prominent ethical theories/lenses • Apply ethical lenses to uses of AI • Create fictional AIs, incorporating imaginative elements and what we've learned about extant machine learning applications • Learn about algorithmic bias as an example of an ethical issue relevant to AI	• Writing prompt: Think of a time when you—or your friends or family—encountered a tricky situation with technology where you had to choose what was right to do. What happened? Was it easy or hard to know what the right decision was? • Intro to ethics and ethical lenses • Watch and discuss (whole-group) self-driving car dilemma video: https://youtu.be/ixIoDYVfKA0 • Science fiction writing activity in groups
Week 5 April 20	Topic(s): • (How) can we explain how AI works? • How can we use ethical lenses to think about AI, continued: how can we apply ethical lenses to contemporary AI issues? Objectives: • Define and describe deepfakes and think about how this technology can have beneficial and harmful applications • Work with Markkula ethical decision making framework, applying ethical lenses to AI ethics scenarios (from real-world, contemporary issues)	• Writing prompt: 1.What's something you're glad you learned so far? 2. What's something you're confused about or still wondering about? 3. What's something you want to learn? • Watch and discuss Crash Course Algorithmic Bias video: https://youtu.be/gV0_raKR2UQ • Learn about explainable AI (slide) • Watch 60 Minutes deepfakes video: https://youtu.be/Yb1GCjmw8_8 and brainstorm 3 ways this technology could be used for good and 3 ways it could be used to cause harm • Work through ethical dilemma scenarios using the Markkula

		framework and handout with stakeholder table
Week 6 April 27	Topic(s): ● AI for social good: How can AI be used for education and preservation, and what ethical issues does this entail? Objectives: ● Explore an application of AI for holocaust education, thinking about the ethical ramifications ● Interact with an AI Holocaust survivor and discuss the experience in small groups and as a class	Day 1 ● Writing prompt: After watching the video your teacher showed you yesterday, how do you feel about police using AI? Who benefits from this technology, and who could be harmed by it? Are the benefits worth the drawbacks? Do you think this is an ethical use of AI? Why or why not? ● Watch (first 8 mins) and discuss video on AI for holocaust education: https://youtu.be/D9tZnC4NGNg ● Interact with AI holocaust survivor Pinchas Gutter:https://iwitness.usc.edu/dit/pinchas ● Table group discussions on experience with Mr. Gutter (questions on handout) ● Whole-group share out and discussion ● Work on science fiction writing in groups (finish stories and submit for homework)
Week 7 May 4	Topic(s): ● Technical perspectives: What is machine learning and how does it work? Objectives: ● Explore AI as a technical phenomenon (machine learning, neural networks, GANs) ● Embody how neural networks learn by playing Neural Network Game (MIT)	● Share out science fiction AIs ● Slides on machine learning (from Brian Broll), supervised learning, neural networks ● Neural network game (from MIT) ● GAN or Not Game (from MIT) ● Writing prompt (exit ticket): what connections to ethics can you make? What harm and good do you think can come from the GANs we've seen today?
Week 8 May 11	Topic(s): ● Sci fi/speculative perspectives: How is AI portrayed in popular	● Watch *Ron's Gone Wrong* movie ● Fishbowl discussion of *Ron's Gone Wrong*

	media, and (how) is that significant? Objectives: ● Make connections between science fiction and our own lives, thinking about how science fiction can be a tool for social commentary and imagining alternative futures ● Use fishbowl discussion protocol to discuss science fiction film *Ron's Gone Wrong*	● Give out superlative awards for science fiction stories ● Assign "QCQC" (quotes, comments, questions, connections) handout for homework
Week 9 May 18	Topic(s): ● AI & environmental impact: how can AI harm and/or help the environment? ● Understanding reports about AI: how can we interpret information about AI when we encounter it? ● Connecting to current events: How does AI affect our world and how is it reported in the news? Objectives: ● Define and discuss key issues in sustainable/green AI ● Interpret news stories about AI, identifying the perspective of the author and the ethical issue they discuss ● Conduct one-on-one interviews with students who volunteer	● Introduce AI & environmental impact ● AI in the news activity: introduce media bias chart, then read and answer questions on news articles in small groups ● Work on review sheet with key terms and topics ● One-on-one interviews with students who volunteer
Week 10 May 24, interviews, no whole class meeting	Objectives: ● Conduct remaining one-on-one interviews with volunteering students	● One-on-one interviews with students who volunteer

*Where indicated, I've used materials from other researchers, including from The DAILy Curriculum for Middle School Students, created by the MIT Media Lab Personal Robots Group and the MIT STEP Lab (see https://raise.mit.edu/daily/index.html), Blakely Payne's AI + Ethics Curriculum for Middle School (https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/), and slides from Dr. Brian Broll.

Appendix B: 8th Grade Curriculum Overview

AI & Ethics Curriculum Overview
8th Grade

Sessions	Objectives/Topic Overview	Activities
Week 1 March 22 & 24	Topic(s): ● What is AI? ● How does AI work? Objectives: ● explore AI from a variety of perspectives, identifying many examples in our world ● develop working definitions of AI for the unit ● consider connections between fictional and currently implemented AIs ● Explore AI from a computational perspective, training an AI model to (begin to) understand how it works, what breaks it	Day 1 ● Introduction to Sarah and research project ● Overview of unit ● Assent process ● AI perspective scavenger hunt on jamboard—find examples of AIs from different perspectives ● Define AI at tables and present definitions on chart paper Day 2 ● Writing prompt: What do you think you need to know to be about to make an AI? What do you think you need to know about how it works? ● AI Bingo (from MIT materials)* ● Watch and discuss video: Machine Learning: Living in the Age of AI | A WIRED Film at https://www.youtube.com/watch?v=ZJixNvx9BAc&t=3s ● Teachable Machine activity
Week 2 March 29 & 31	Topic(s): ● Where is AI? ● Where is your data? Objectives: ● identify/locate everyday AI, surveillance & data-gathering mechanisms ● Develop ideas about how (students think) data should and should not be used/collected	Day 1 ● Writing prompt: If you could solve any problem using technology, what would you solve and how? ● Defining data ● Algorithmic bias video (Joy Buolamwini): https://youtu.be/TWWsW1w-BVo ● AI Bingo game (MIT) ● UNICEF video and discussion: https://youtu.be/MBEPjFTuTPw Day 2 ● Read and mark up UNICEF manifesto excerpt ● Brainstorm where data is and who

		is collecting it at table groups, then discuss whole group ● Write 3 rules/policies for data ● Watch NYT Facebook and Washington Post's TikTok (first 5 mins) videos ○ https://youtu.be/JAO_3EvD3DY ○ https://youtu.be/nfczi2cI6Cs ● Discuss videos and make benefit/risk tables on chart paper in table groups: What are the benefits or the potential benefits that you can imagine of using data in the way that we saw in these videos? ● Hand out AI Audit to complete as homework
Week 3 April 5 & 7	Topic(s): ● Where is AI? ● Where is your data? ● How does AI classify you and others? ● What are the connections between "real" and fictional/imagined AIs? Objectives: ● Complete personal data and AI use audits, identifying where and when AI is using our data/we use AI in our daily lives ● Create fictional AIs, incorporating imaginative elements and what we've learned about extant machine learning applications	Day 1 ● Writing prompt: What are the most important aspects of your identity? What are the top 5 things that describe you? Now…how do you think you are categorized by the systems that target you (e.g., your social media feed, the ads you see on google)? Who do advertisers think you are? What differences do you see between how you see yourself and how an AI might see you? What does your data capture about you and what does it miss? ● Read and mark up day in the life of your data handout from Apple ● Investigating Google Ads settings—how does Google categorize me? Day 2 ● Writing prompt: What are the biggest issues with technology today? ● Whole-group reflection on audits ● Watch and discuss 60 minutes video on MIT lab: https://youtu.be/QTXrez8u_J0 ● Speculative fiction writing

Week 4 April 12 & 14	Topic(s): ● What are ethics and ethical lenses? ● How can we use ethical lenses to think about AI? Objectives: ● Define ethics and understand several prominent ethical theories/lenses ● Apply ethical lenses to uses of AI ● Introduce final projects ● Learn about algorithmic bias as an example of an ethical issue relevant to AI ● Review key terms and create review materials (30 second digital stories)	Day 1 ● Writing prompt: Do you think fictional AI is important for understanding the AIs we use or will use? Why/why not/how? Come up with an example that supports your opinion. ● Intro to ethics and ethical lenses ● Watch and discuss self-driving car dilemma video: https://youtu.be/ixIoDYVfKA0 ● Fill out topic interest survey for final projects ● Writing prompt (exit ticket): What tricky situations can you think of relating to technology? Think about a time when you or someone you know had a dilemma (situation with difficult choices) when using technology. What happened? Day 2 ● Review ethical theories ● Watch and discuss Crash Course Algorithmic Bias video: https://youtu.be/gV0_raKR2UQ ● Discuss AI-specific ethical issues ● Apply ethical lenses to example from video ● 30-second digital stories: pick a term, write a script, and create a 30-second video to explain it to the 7th graders (use planning handout)
Week 5 April 19 (no meeting on Thursday this week)	Topic(s): ● (How) can we explain how AI works? ● Digging deeper into ethics: what are meta-ethics? ● How can we use ethical lenses to think about AI, continued: how can we apply ethical lenses to contemporary AI issues? Objectives: ● Define meta-ethics and explore meta-ethical theories ● Work with Markkula ethical	Day 1 ● Writing prompt: 1.What's something you're glad you learned so far? 2. What's something you're confused about or still wondering about? 3. What's something you want to learn? ● Watch 30 second stories from last time ● Learn about explainable AI (slide) ● Define meta-ethics ● Discuss meta-ethics whole-group ● Work through ethical dilemma scenarios using the Markkula

	decision making framework, applying ethical lenses to AI ethics scenarios (from real-world, contemporary issues)	framework Day 2 (no meeting)
Week 6 April 26 (no meeting on Thursday this week)	Topic(s): ● AI for social good: How can AI be used for education and preservation, and what ethical issues does this entail? Objectives: ● Explore an application of AI for holocaust education, thinking about the ethical ramifications ● Interact with an AI Holocaust survivor and discuss the experience in small groups and as a class ● Continue work on final projects	Day 1 ● Writing prompt: How can we prevent unethical uses/applications of AI? Give examples. How should we address harmful outcomes or uses of AI when they happen? ● Watch (first 8 mins) and discuss video on AI for holocaust education: https://youtu.be/D9tZnC4NGNg ● Interact with AI holocaust survivor Pinchas Gutter:https://iwitness.usc.edu/dit/pinchas ● Table group discussions on experience with Mr. Gutter (questions on handout) ● Whole-group share out and discussion ● Work on final projects Day 2 (no meeting)
Week 7 May 3 & 5	Topic(s): ● Technical perspectives: What is machine learning and how does it work? ● Sci fi/speculative perspectives: How is AI portrayed in popular media, and (how) is that significant? Objectives: ● Pair exploration of AI as a technical phenomenon (machine learning, neural networks, GANs) and as a speculative phenomenon (in science fiction film) ● Play neural network game (MIT) to embody how neural networks learn	Day 1 ● Watch first part of *Ron's Gone Wrong* movie ● Writing prompt: Predict: what do you think is going to happen to Ron? Write down notes about movie's commentary on tech and society ● Slides on machine learning (from Brian Broll), supervised learning, neural networks ● Neural network game (from MIT) Day 2 ● Watch rest of *Ron's Gone Wrong* movie ● Complete "QCQC" handout, writing important quotes,

	● Make connections between science fiction and our own lives, thinking about how science fiction can be a tool for social commentary and imagining alternative futures ● Use fishbowl discussion protocol to discuss science fiction film *Ron's Gone Wrong*	comments, questions, and connections to the movie ● Fishbowl discussion of *Ron's Gone Wrong* ● Introduce Generative Adversarial Networks (GANs) with slides and game from MIT ● Hand out final project part 2 sheet
Week 8 May 10 (no meeting on Thursday this week)	Topic(s): ● AI & environmental impact: how can AI harm and/or help the environment? ● Understanding reports about AI: how can we interpret information about AI when we encounter it? Objectives: ● Define and discuss key issues in sustainable/green AI ● Interpret graphs depicting trends in AI, identifying the major argument of the graph and how the visuals support it	Day 1 ● Consult with students about final projects during study hall ● Discuss meaning of critical questioning and why it's important ● Introduce AI & environmental impact ● Interpreting information about AI: graph activity with Stanford's AI Index ● Hand out part final project part 3 sheet Day 2 (no meeting)
Week 9 May 17 (individual interviews, no whole class activities on Thursday)	Topic(s): ● Connecting to current events: How does AI affect our world and how is it reported in the news? Objectives: ● Conduct one-on-one interviews with students who volunteer	Day 1 ● Work session on final projects ● Chart paper discussion of quotes from *AI Ethics* text ● Table group discussions of AI's connection to online radicalization, Buffalo shooting ● News article activity—choose recent news article to read with partner and answer questions on handout, present to whole group Day 2 ● Answer questions/help with final projects ● One-on-one interviews with students who volunteer
Week 10 May 24	Topic(s): ● Share and celebrate work Objectives: ● Conduct remaining one-on-	Day 1 (final visit) ● Writing prompt: Design an activity for teaching someone about your project topic.

	one interviews with volunteering students ● Students present final projects to the class	● One-on-one interviews with students who volunteer ● Final project presentations (not mandatory but highly encouraged)

*Where indicated, I've used materials from other researchers, including from The DAILy Curriculum for Middle School Students, created by the MIT Media Lab Personal Robots Group and the MIT STEP Lab (see https://raise.mit.edu/daily/index.html), Blakely Payne's AI + Ethics Curriculum for Middle School (https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/), and slides from Dr. Brian Broll.

Appendix C: Design Conjectures

Theoretical and Design Conjectures

The high-level design conjecture that guides this project is: *through engaging with AI in peer discussions, direct instruction, and individual and collaborative interdisciplinary activities, students will develop multifaceted understandings of AI, identify and critique AIs operating in their daily lives, and develop ideas about ethical and agentive engagement with them*. This high-level design conjecture is supported by three related theoretical conjectures that address the nature of AI, the role of meta-ethical engagement with AI, and the value of posthumanist agency for understanding our entanglements with AIs.

Theoretical Conjecture 1: AI as a Complex Phenomenon

Exploring AI through interdisciplinary lenses and methods will help students see AI as a complex technical, social, and ideological phenomenon, and build on ways they already do so. This conjecture theorizes about AI itself, building a definition that extends beyond the technical, to honor the myriad ways students encounter AI in computing devices, but also in film, in popular discourse, and in marketing. I also hypothesize that using interdisciplinary approaches (elaborated in design conjecture 3) contributes to the development of this sort of broad, complex understanding of AI.

Furthermore, discussions about AI's complex presence in our lives will be designed to help students locate or "see" where AIs are operative, develop some understanding of what AIs are and how they operate, and critically evaluate the presence and operation of the AI.

Theoretical Conjecture 2: Meta-Ethics

Directly discussing theories and applications of ethics (meta-ethics) will allow students to explore complex current and hypothetical entanglements with AIs and investigate non-normative perspectives.

In this conjecture, I posit that giving students opportunities to engage with meta-ethics will allow them to critique normative perspectives and to imagine counterfactual or future scenarios.

Theoretical Conjecture 3: Entangled Agencies

Understanding AIs and humans as entangled in agentive webs will allow for new insights into our potential for, and constraints on, action and change.

My second theoretical conjecture re-theorizes agency as a result of human and non-human entanglements. By shifting ideas of agency away from narrowly humanist conceptions of student choice, I am anticipating that my students will see new possibilities for envisioning the possible actions they and the AIs they are enmeshed with can take.

Following from these theoretical conjectures, I have formulated a set of design conjectures expressing my hypotheses about how design elements might contribute to certain student outcomes. Each design conjecture features a pedagogical technique, practice, or activity design that builds on previous research and is to be tested in further iterations of the design. It is my hope that the activities and practices outlined in the design conjectures might contribute to multiple theoretical conjectures, so it is not a direct one-to-one correspondence between theoretical and design conjecture.

Design Conjecture 1: AI/Data Audit

AI/Data audits can support students in looking for and recognizing where and how AI is operative in their lives, analyzing the power relationships involved in how we use AI/AI uses us, and brainstorming/identifying possible actions they might take to change their use of/by the AI if they deem it necessary.

The AI/Data Audit is a tool I drafted for use with my undergraduate seminar students. In the audit, I asked students to notice and record where they encounter or use AIs, and where they generate digital data, over a specified time period. In the audit's supplied chart, students

were encouraged to record these interactions, try to figure out who operates the AI and why, critically evaluate the benefits and risks of using/being used by the AI, and finally brainstorm potential actions they might take, should they want to change how they use/are used by AIs. After the audit was complete, students wrote a brief reflection on the experience of doing the audit and on anything that surprised them.

I will use a revised version of this tool in my next iteration of data collection. Although students will create a final product, an important part of this activity is also the scaffolding that leads up to completing the audit. For example, we will discuss how their actions might be constrained or enabled by entanglement with AIs.

Design Conjecture 2: Meta-Ethical Analysis Strategies

Analyzing extant codes of ethics and creating, analyzing, and discussing their own codes of ethics for AIs will support development of students' meta-ethical analysis skills and practices.

This conjecture outlines three strategies for meta-ethical engagement: analyzing codes of ethics that exist (for example, in fields like journalism and medicine), creating a personal code of ethics for using/being used by AIs, and discussing codes of ethics with others. I hypothesize that asking students to identify and discuss engagements with AIs as ethically significant, examine their own ethical beliefs, and discuss them with others (meta-ethical engagement) will deepen their ideas about and discussions of agency, potentially shifting student thought and/or actions around their/others' uses of/by AIs. Dialogue in various configurations (whole class, small group, pairs; unstructured and structured around specific questions) will be an important part of this design. Hearing other people's questions, beliefs, and opinions, I posit, will contribute to individual and collective development of ethical and meta-ethical analytical practices.

Design Conjecture 3: Interdisciplinary Activities

Both technical exercises (e.g., tinkering with AI design via coding) and fiction reading/viewing and writing are valuable disciplinary inroads into understanding AI, and using both will deepen engagement with AI.

This includes activities like reading, analyzing, and writing fictional texts, alongside activities that allow students to understand and tinker with the technical operation of AIs. An example of a technical exercise (borrowed from Vanderbilt's Text Mining Working Group) might be to use blocks-based programming to play with an AI that analyzes the sentiment (positive, negative, neutral) of text you write. Writing a piece of fiction about an AI is an example of an activity that need not be technical. Writing a piece of fiction *with* an AI could be an activity that allows for both technical tinkering and creative engagement.

In addition to the conjectures listed above, I want to add an additional "design" conjecture to note that I will still try to be attuned to ways that students draw on resources and experiences outside of my design, and outside of the classroom, to learn about AIs and develop critical and ethical responses to them. This un-design conjecture states that no design could or perhaps should be the sole locus of learning, so I will build in opportunities for students to connect to their life and world outside of the designed activities and materials; soliciting student input about ways they find, understand, and critique AIs and explore their agency outside of the designed activities will allow me to explore new ideas extraneous to the (current) design and to build on student experience and expertise.

Appendix D: Teaching Activities

AI & Data Audit

AI and Data Audit

Directions:

- Choose a time frame to record your data/AI uses. This can be 24 hours, or the highlights of a few days.
- Make a note of when you generate data that is collected by technology or you think you're using AI. Please exclude anything that is more personal than you'd like to share! Notice things like when you use
    - Your phone--location services, apps, social media, texting
    - Your computer--google searching, email, games, school systems
    - A credit card
    - A rewards card, like for a grocery store or Starbucks
    - A home assistant like Siri or Alexa, or smart appliances like Ring doorbells
- Your product will be a filled out table (see below) and a reflection on the activity

Name: _____

What/Where/When Describe the activity, time, and place here. What data is being collected?	**Who & Why** Who is collecting the data/using the AI, and why are they doing so? (In some cases, you might not know the answer to this!)	**Critical Evaluation** How does this benefit you? Does this concern you? Why or why not? What security protections or issues could you envision?	**Possible Actions** Do you approve of this use of your data? How might you opt out or control your data/use of/by AI here?

Reflection:

In the reflection, consider

- Did you see what you expected to see?
- What information was easy to gather, and what was difficult to see or find out?
- What are your general thoughts about your everyday encounters with AI? Your data privacy and security? Your options for taking action?

Ron's Gone Wrong QCQC

This template is borrowed from Dr. Melanie Hundley, and I have adapted it for use with the film we watched.

QCQC Film Response

Name_____

Quote	Comment
Write down a key quote from the movie that you want to discuss or that you think is from an important moment. Explain why you picked this quote.	What did you think about the movie and its message(s)?
Questions	**Connections**
Write at least one discussion question related to the movie. What are you still wondering about? What question would you like to pose to the group, or would you like to ask the filmmakers?	How do the themes of this movie connect to your own life and/or things you hear in the news?

Final Project Instructions Parts 1, 2, & 3

AI & Ethics Project
Part 1

About the project:

- Objectives: explore an area of interest, develop expertise, effectively communicate what you learn to others, apply learning about AI and ethics, work in supportive groups with peers
- The project has 3 parts:
    - 1. Background information/research
    - 2. Application of ethical thinking/decision making framework
    - 3. Speculative fiction writing/film/art
- You will create your own, individual products for each part, but you will work in supportive groups to provide feedback and workshop ideas.

Today, you'll investigate your area of interest, narrow your topic, choose the format of your project, and start research.

Step 1: Investigate your area of interest.
In order to pick your more specific topic, you'll need to get a sense of the major issues, advancements, and tools in your chosen area of interest. You can search the web for reliable information (check with us if you have questions about whether something is reliable!). For example, you might want to go to a trusted news source and search for the topic on the site. A resource on AI trends in general is https://aiindex.stanford.edu/report/.

Step 2: Narrow your topic.
After you spend a little time reading about your area of interest, identify a more specific topic within it. This could be a specific research project, a use of a specific AI tool/product, or a particular ethical issue within this field.(If you come across an AI or ethical issue you're interested in that's not in your initial area of interest, feel free to switch to that at this point!)

Write your topic in the google doc once you choose it.

Step 3: Choose your format.
This can be any of the options below. After you choose, find at least one really good example of this format to use as a guide.

- An ebook (e.g., using something like Book Creator)
- A slide presentation (google slides, powerpoint)
- A video (using animoto or another site/software)
- A website
- A public service announcement
- Infographic
- Podcast episode
- Another format of your choosing, approved by [your teacher]

Write your format choice in this google doc once you choose it.

Step 4: Share your topic and format choices with your group members.

Check in with your groupmates. Does everyone have a specific application of AI they are researching? Do you have any particular expertise you might be able to offer with the format they've chosen? Do you know of any cool examples of that format they might want to use for inspiration? Did you come across any resources in your general search that might help your groupmate?

Step 5: Start your research.
Now, find out as much as you can about your topic. Answer the questions below (adjust as needed if your topic is a research project rather than an AI tool). Gather and organize your sources in the source chart, and start thinking about how you might incorporate the information into your chosen format.

- Who made the AI/research project and why?
- What does it do?
- Who uses it?
- How does it work? What data does it need to work, and how does it collect that data?
- What ethical issues have been reported about it?

Source Link	Author/Site Description	Short Description
Example https://www.vice.com/en/article/pa7dj9/flawed-algorithms-are-grading-millions-of-students-essays	Todd Feathers, journalist for Vice, a big media company	Vice news article from 2019 exposing widespread use of grading AIs, especially in standardized testing. Has links to lots of source material.

AI & Ethics Project
Part 2: Exploring Ethical Issues

In this section of the project, you will explore the ethical issues related to your topic.

1. What are the major ethical issues?
    a. Research: find out if anyone has written about ethical issues surrounding your topic. What issues have you noticed while researching your topic where good/bad, right/wrong uses of the technology have come up?
    b. Write: make a list of the ethical issues you can find (noting where you found them) or can imagine. Think about
        i. How can this technology be used for the most good?
        ii. How could it cause harm?

2. Describe an ethical dilemma.
    a. On the handout, write a brief description of an ethical dilemma that could arise or has arisen from using the technology.
    b. Use the framework handout to think through the dilemma from various ethical perspectives.
3. Write a code of ethics.
    a. Below, write at least three rules for using the technology ethically. Think about the things we've talked about in class, like transparency, privacy, consent, algorithmic bias, responsibility/authorship.

AI & Ethics Project
Part 3: Imagining the Future

This is your chance to flex your creative muscles! Pick a time in the future. This could be 5 years from now or more than 500. Think about what the world will look like then. How will your AI be used then? How do you think the technology will have developed/changed? Jot down your thoughts here:

Imagine how things could go wrong in the future with your AI. Could using (or not using) your AI lead to suffering or injustice? Describe a dystopian future here:

It's the year_____ and things are not looking good.

Imagine how your AI might be used to make the world a better place. How might your AI benefit humanity/the planet/other life forms? Describe your best-case, utopian future here:

It's the year_____and things couldn't be better.

30-Second Stories

This template is adapted from Bridget Dalton's original activity.

Your goals:

- To teach your classmates about your term by showing, not telling, about it in a 30ish second film.

Steps:

1. Research/remind yourself what your term means. Use your notes and search for reliable information online (ask me or your teacher if you are unsure about whether a site is reliable!). Write a definition, some synonyms, and an example below.

Your term:
Definition(s):
Synonyms:
Example:

2. Write your term—make it large and legible—on a piece of paper. You'll show this at the end.
3. Think about how you might SHOW what that term means. Create a scene where you might use the term and/or show what it means.
4. Write a script for your story. Remember to show your term at the end.

Script:

5. Collect any props you might need for the story and a camera.
6. Practice your skit, then film!
7. Review your film and re-film if needed or upload finished video to Google Classroom.