# In the Black Mirror: Youth Investigations Into Artificial Intelligence

In the Black Mirror

Youth Investigations Into Artificial Intelligence

CLIFFORD H. LEE

Mills College, School of Education & YR Media, clifflee@mills.edu

NIMAH GOBIR

YR Media, nimahgobir@gmail.com

ALEX GURN

Rockman et al, alex@rockman.com

ELISABETH SOEP

YR Media, lissa.soep@yrmedia.org

Over the past two decades, innovations powered by artificial intelligence (AI) have extended into nearly all facets of human experience. Our ethnographic research suggests that while young people sense they can't "trust" AI, many are not sure how it works or how much control they have over its growing role in their lives. In this study, we attempt to answer the following questions: 1) What can we learn about young people's understandings of AI when they produce media with and about it? 2) What are the design features of an ethics-centered pedagogy that promotes STEM engagement via AI? To answer these questions, we co-developed and documented three projects at YR Media, a national network of youth journalists and artists who create multimedia for public distribution. Participants are predominantly youth of color and those contending with economic and other barriers to full participation in STEM fields. Findings showed that by creating a learning ecology that centered the cultures and experiences of its learners while leveraging familiar tools for critical analysis, youth deepened their understanding of AI. Our study also showed that providing opportunities for youth to produce ethics-centered Interactive stories interrogating invisibilized AI functionalities, and to release those stories to the public, empowered them to creatively express their understandings and apprehensions about AI.

**CCS Concepts** • Social and Professional topics • Professional topics • Computing education • Computing education programs • Computer science education

**Additional Keywords and Phrases**: Critical pedagogy, artificial intelligence, ethics-centered, engagement, agency, computational thinking, machine learning, media

## 1 INTRODUCTION

Over the past two decades, innovations powered by machine learning have extended into nearly all facets of human life. From at-home "smart" technologies for security and comfort; to mobile devices and facial recognition tools that track movements and behaviors; to algorithms that determine social media consumption and marketplace identities, artificial intelligence (AI) has become ubiquitous for most of us who are "connected" through various digital networks. Just to participate in daily practices, we consent to corporate terms and conditions through absentminded clicks of "I agree," hardly registering the ethical decisions or implications of the permissions we grant and data we share with every act of assent. Even more problematically, the AI of our quotidian technologies can operate non-transparently and only come to light when an accident results in headlines exposing corporate misconduct.


/2022/1-ART1 $15.00
http://dx.doi.org/10.1145/3484495

Through our own surveys with youth (ages 14-24), we learned that 77% get most of their information about AI from popular culture, such as science-fiction, movies, TV, and video games. Our research suggests that while young people sense they can't "trust" AI, many are not sure how it works or how much control they have over its growing role in their lives. This study seeks to understand how young people underrepresented in STEM make meaning of the role of AI in their lives and society and how their relationship to the technology evolves when they create their own AI-based tools and media. This research also analyzes the curriculum and pedagogy behind three ethics-centered AI learning activities housed within an after-school multimedia production organization.

At YR Media, a national network of youth journalists and artists, we create content on multiple platforms in numerous media forms (audio, video, textual, multimodal, Interactive, and live performances). Ten years ago, we established Interactive, a department within the organization's newsroom where young people produce interactive stories combining journalism, design, data, and code. The three examples outlined in this study examine the learning processes and ecology involving three cohorts of youth, each one varying between five to nine young people. A few participants were in multiple cohorts. In total, sixteen producers, aged fifteen to nineteen, who are predominantly youth of color and those contending with economic and other barriers to full participation in STEM fields engaged in this research study.

By creating poems using their phones' autocomplete feature; reformulating Spotify's algorithm rating pop songs' danceability; and investigating the preponderance of facial recognition and surveillance technologies, Interactive youth deepened their understanding of AI while challenging, disrupting, and educating others about the invisibilized AI features operating on tools that hundreds of millions of users engage with every day. Youth producers shared their insights through interactive stories the team designed, developed and disseminated to mass audiences online and via social media. Other recent projects from the Interactive department of YR Media include interactive maps highlighting the impact of gentrification through audio and video storytelling [1]; infographics that demonstrate gender biases in school dress code policies [2]; and curation of social media posts from students from every state in the U.S. whose campuses closed due to Covid-19 [3].

The Interactive department prepares young people to harness new digital tools and journalistic methods in co-creating digital media that addresses issues impacting youth communities. The creative process operates at "the intersection of engineering and computational thinking on the one hand, and narrative production and critical pedagogy on the other" [1]. Our ethnographic study centered participant observation, as some of the authors of this paper helped to design and facilitate the learning experiences we describe here and to edit the projects we highlight. We collected audio recordings of moment-to-moment interactions in class, end-of-session focus group interviews, and analyses of youth-generated artifacts within the learning environment over the course of two and a half years. Our research team used a grounded theory approach to code, reduce, and analyze the data to generate themes according to our Critical Computational Expression conceptual framework. Through the study, we demonstrated the potential for computer science (CS) education to embed computational thinking inside a youth-driven newsroom where content is tethered to truth, aimed towards justice, shared with the public, and brought to life through creative expression [4].

## 2 ESTABLISHING CONTEXTS FOR EQUITY AND JUSTICE IN COMPUTING EDUCATION

### 2.1 "Reframing 'Achievement Gaps'" [5]

For more than a century, the U.S. educational system has routinely viewed technology as a resource to counter educational inequality [6, 7]. And yet, despite the billions of dollars invested in the latest digital tools, little has changed in the educational outcomes of historically marginalized students [8]. One has to look no further than the billion-dollar iPad rollout by the Los Angeles Unified School District and its subsequent failings as evidence of the ineffectiveness of technology for technology's sake. Students hacked the security within weeks of its debut, there was insufficient training for teachers on implementation, and the program generated uneven indicators of improved student learning outcomes [9].

To understand the problem with turning to technology to address so-called "achievement gaps," itself a deeply flawed construction, one must take into consideration the historical antecedents and contemporary contexts that produced these differences, which Ladson-Billings aptly frames as an "educational debt" [5]. She refocuses our understanding of educational inequality as the continued legacies of systemic, intergenerational, and institutionalized racist, classist, and sexist policies and practices that exert persistent influences on student performance. Far too often, students are expected to adopt reigning epistemological traditions, ideologies, and practices in order to be successful in mainstream educational spaces without acknowledgment of the "community cultural wealth" they bring to learning or their "rightful presence" in STEM spaces in particular [10, 11, 12].

Ladson-Billings' argument acknowledges sociocultural learning theorists' view that learning is mediated by a learner's engagement and participation through cultural interactions with people and tools [13]. Hence, meaning-making occurs jointly between teachers and students, through their respective cultural experiences, practices, and epistemological backgrounds [14, 15, 16]. Thus, the implementation of a one-size-fits-all technological solution will continue to leave historically marginalized students on the educational periphery [17]. That is, until the learning ecosystem (teachers, curriculum, pedagogy, schools, districts, administrators, policies, rules, and tools) reconfigures itself to recognize the diverse cultures of its learners

### 2.2 Centering Equity and Empowerment with Culturally Responsive STEM Education

A justice-driven [18] and equity-centered approach [19] squarely places inequities as a key unit of analysis for student learning. In the process of developing critical consciousness related to our shared social world, students demystify, analyze, interrupt, and challenge educational disparities perpetuated through technology. They develop agency through their engagement with models and explanations to understand processes of marginalization and take action towards social transformation, in some instances speaking truth to power using the same technologies that fail to recognize or serve them [19, 1]. Until students are supported and taken seriously as next-generation justice-oriented creators, they can be reduced to users of products dictated by a narrow strata of mostly white, cisgendered, heterosexual males monitoring and shaping the needs, wants, and desires of huge swaths of the human population.

Examples of technology and culturally-responsive STEM education grounded in indigenous connections [20, 21] offer a counternarrative that is sorely needed. Another critical aspect is the role of teaching, pedagogy, and sustained relationships within classrooms. While we do not discount the importance of technological access, particularly as seen through the perpetuation of inequities in distance learning during the Covid-19 pandemic, we have also seen countless examples where access is addressed, yet results are mixed [22, 23, 24, 25]. Simply mandating a CS course or coding bootcamp in an under-resourced school does not ensure a culturally-responsive approach, nor does it fundamentally address the institutional barriers that students and their communities face.

### 2.3 Examining AI's Potential to Reflect and Exacerbate Bias

As AI becomes more pervasive and computers take on tasks typically performed by humans, the technology is reshaping people's social interactions and expectations [26, 27]. Recent advances in computational processing and machine learning have enabled companies to integrate AI into a broad range of products and services, for instance, monitoring credit card activity to detect potential fraud, using facial recognition to unlock one's phone or identify faces on one's camera roll, predicting text to suggest the next words in a sentence, or selecting content to display in one's social media news feed [28].

Proponents of AI maintain that its applications hold the potential to enhance productivity, to tackle society's most pressing problems, and to advance social and economic opportunities [29]. AI is being deployed for disease prevention and treatment, to identify and address online abuse, and to intervene in the climate crisis, to name just a few promising applications [30]. That said, critics warn that if implemented without a robust ethical framework, AI threatens to reify existing biases, deepen social inequities, and undermine privacy in the digital age [31, 32, 33]. Since the algorithms powering machine learning reflect the assumptions of their developers and the parameters of the data sets employed, there is a need for the developers and users of AI to critically reflect on the nature of the technology, its uses, and the extent to which it adheres to defensible principles [34].

### 2.4 Accountability in AI Product Design

Despite AI's growing influence on our digital environments and experiences, the public is often uninformed or unaware that AI is embedded in the technologies they use regularly [35]. One recent exploratory study of youth and their parents about children's media use found that AI is understood through the lens of popular culture [36]. Participants did not independently recognize or pay attention to the presence of AI technology in the digital devices or services they accessed. Moreover, youth and parents did not identify potential related issues, such as data privacy or information sharing. In contrast to the common assumption that contemporary youth are naturally adept "digital natives", many young people simply do not recognize the issues posed by the technology they use every day.

Both the design of the technology itself and digital literacy education play roles in young people's misunderstandings of the role of AI in culture and society. Machine learning algorithms, designed to discover patterns in large data sets without explicitly programmed instructions, are commonly referred to as "black boxes" as a result of their opaque inner workings that prevent users from knowing how or why a neural network makes specific decisions. Sometimes the designers themselves cannot explain the results

produced by the algorithms. Thus, the computational mechanisms shaping one's social media newsfeed or music app recommendations remain hidden from the user interface or publicly-available information, often protected as intellectual property [37]. The fact that most machine learning processes are black-boxed creates a key challenge for AI education. When young people lack opportunities for direct exploration of the technologies they use, their "interaction with black-boxed processes may lead to the development of inaccurate or oversimplified mental models" [38]. Without intentional learning activities directed at uncovering how AI works, youth may not truly come to understand the technology.

Recognizing that algorithmic black boxes can exert powerful influences over institutional decision-making and reflect the biases of the programmers or the training data, AI scholars and legal experts increasingly have called for greater transparency and accountability of algorithms [39]. Outside the United States, the European Commission has underscored that citizens have a "right to explanation" of decisions impacting them, calling for research on the explainability of AI in which the results or decisions reached by the technology can be understood by its users and designers [40]. Recent antitrust efforts and scrutiny of big tech's role advancing disinformation suggest a growing sense of urgency to hold these companies accountable in the United States as well.

However, given the widespread lack of public understanding or clarity about when and how algorithms operate or the sorts of influence they can have, even increased transparency may not lead to adequate knowledge among technology users [41]. At a minimum, there is a need to incorporate awareness of AI and machine learning algorithms into information and digital media literacy in order to enhance "autonomy to the user and to support the critical use of digital media and understanding pertaining to how different elements of the technological and social system affect and shape the public sphere" [43]. This study contributes to the small but growing body of research on how young people perceive and make sense of AI's presence in their lives and promising educational practices that develop students' understanding of the technological or socio-cultural dimensions of AI [42, 43].

### 2.5 Building Critical Consumers and Producers with AI Education

Recent educational research has examined ways to build AI literacy among stakeholders without specialized STEM training, developing educator guides on how to teach CS principles and practices related to AI [44, 45] Notably, the Association for the Advancement of Artificial Intelligence (AAAI) and the Computer Science Teachers Association (CSTA) collaborated to develop a set of learning guidelines for K-12 classrooms to teach students about AI [46]. The group identified five "Big Ideas" in AI that they believe students should know: 1) Computers perceive the world using sensors, 2) Agents maintain models/representations of the world and use them for reasoning, 3) Computers can learn from data, 4) Making agents interact with humans is a substantial challenge for AI developers; and 5) AI applications can impact society in both positive and negative ways [46].

As society becomes more infused with and dependent on AI and other "invisible technologies" [47], youth need to develop awareness and understanding of the complex nature and roles of AI and how it shapes their lives. The five "Big ideas in AI" offer a framework to help introduce students to the core concepts necessary to become critical users and responsible producers of AI.

Recent research suggests that hands-on learning can help youth to avoid or overcome inaccurate perceptions of AI. For example, Druga et al. (2018) examined parents' and children's assessments of a robot's abilities. Findings suggested that hands-on experience exploring the underlying processes of the robot helped participants develop more sophisticated mental models of the AI powering the robot's steering system [42]. In a related study of youth learning by uncovering black-boxed machine learning, Hitron et al. (2019) found that hands-on experience coupled with real-time feedback and reflection can contribute to accurate understanding of AI concepts [38]. These studies demonstrate that children can learn basic technical concepts of machine learning and iterate through trial and error.

The relevance of the learning context in real-world scenarios plays a key role in youth developing nuanced understanding of AI. To test personalized learning strategies for teaching machine learning concepts, Register and Ko (2020) evaluated the effects of three instructional approaches on students' ability to identify and self-advocate when AI models make mistakes or cause harm [48]. The study found that using learners' own personal data helped contextualize machine learning and allowed learners to better ground their critiques of flawed models in the actual mechanisms of the machine. The intervention intentionally drew on learners' "funds of knowledge: leveraging the learner's already existing knowledge and experience by strategically teaching material revolving around the learner's culture, situated knowledge, and relationships" [48].

These related research studies offer insights for our own research as well as opportunities to reinterpret or extend the findings of these prior works. In particular, our research explores AI education in a learning community devoted to examining social issues through media making, with a focus on adolescents — teens and young adults. Positioning youth as collaborators working alongside STEM professionals and their peers, the enduring goal is to create meaningful learning conditions that convert young people's curiosity into active investigation and production that can nurture their sense of purpose and agency in society. Using an iterative, learner-centered approach to design and production, we draw connections to those educational interventions incorporating personally-relevant learning contexts, messy data sets, heuristic learning through trial and error, and real-world applications. At the same time, much of the research on AI education does not explicitly address how current uses of AI sustain or reproduce societal inequalities. Finally, our research provides new insights by highlighting a STEM education environment where young people share products of their learning not only amongst themselves or with their teachers, but also with a broader public through online distribution of their media reaching significant audiences. With an eye to developing the resources and talents of young people historically underrepresented in CS and STEM fields, this paper presents research on curriculum and pedagogy aimed at preparing youth to explore the complex connections between digital technology, social dynamics, and ethical questions in the context of their lives.

## 3 RESEARCH QUESTIONS

If young people are to develop ethically-grounded understandings of AI in practice, they need opportunities to connect computational thinking to civic issues and creative expression. This study advances the development of strategies that promote learning related to AI through collaborative media production. This paper addresses the following research questions:

1. What can we learn about young people's understandings of AI when they produce media with and about it?
2. What are the design features of an ethics-centered pedagogy that promotes STEM engagement via AI?

The first research question sought to understand the nuances and complexities of young people's prior knowledge and experiences related to AI. Namely, we explored where their conceptions come from and how they shift in the process of making AI content through media. YR Media's nearly three decades of experience teaching and producing media with youth has positioned us well to cultivate and nurture learning ecologies to support deep learning on a given topic. The second question sought to uncover the unique qualities of our curriculum and pedagogy that support STEM learner engagement through AI content. We hope our findings contribute to the larger CS and STEM education and research community to strengthen our collective wisdom in promoting STEM engagement, particularly amongst historically marginalized populations.

# 4 CONCEPTUAL FRAMEWORK, THEORETICAL UNDERPINNINGS, AND METHODOLOGY

## 4.1 Conceptual Framework

Critical Computational Expression (CCE) is a theoretical and conceptual framework we have developed that integrates the three distinct traditions of: critical pedagogy, computational thinking, and creative expression. Conceptualized and practiced by Paulo Freire, critical pedagogy promotes non-hierarchical, humanizing teacher-student relationships to engage learners toward critical consciousness and their subsequent liberation by using praxis, or the recursive cycle of action, theory, and reflection [49]. Computational thinking, redefined by Jeanette Wing in 2006, emphasizes the four steps of problem-solving: 1) decomposing a problem (breaking down into smaller parts), 2) recognizing patterns, 3) applying general principles, and 4) designing algorithms [50]. We determined that computational thinking builds on praxis by offering a cognitive and applied model for a problem-solving process in the creation of both digital and non-digital products, while critical pedagogy offers the sociopolitical contextual grounding for its application in a world governed by white supremacy, capitalism favoring the rich and powerful, and patriarchy. Our prioritizing of creative expression draws on approaches to STEAM (science, technology, engineering, arts and math) that refuse to reduce the arts to a mere means for teaching STEM [51]. Learners are invited to notice, observe, reflect on, and represent their feelings, thoughts, and identities through words, movement, visuals and the integrated aesthetics of products they create. We honor traditions of making and expressing the self and culture rooted in marginalized communities that foster habits of healthy risk-taking while aligning with efforts to advance social justice. Also, in very practical terms, young people want their interactive digital media to look and sound great. That won't happen without arts as a throughline in critical and computational thinking and action.

Current research literature and practice contain compelling examples of two of the three tenets of Critical Computational Expression working in conjunction with one another: for example, computational thinking and creative expression as seen in creative code [52, 53] and programmable objects as art [54]; or critical pedagogy and creative expression in public arts projects [55, 56] and youth participatory action research [57, 58, 59]. That said, we have found relatively few examples embodying all three. Coming from

the art practice tradition, critical digital making [60] comes closest to this full-blown triangulation of the tenets, but with a less explicit focus on the role of critical pedagogy in the learning ecology or the dissemination of public products to directly engage in social transformation. Our past studies have demonstrated the power of Critical Computational Literacy [1, 4, 61] as a conceptual and pedagogical framework that supports young people underrepresented in STEM fields in developing their critical consciousness and STEM learning by producing interactive digital tools for public dissemination, but that prior research did not fully develop creative expression as a key factor in youth engagement and social impact. Civic Imagination [62], based on the idea that social change begins through a process of collective imagining and expression enabled by digital tools and new forms of cultural participation has been an instructive conceptual resource guiding our approach as well, and to it we bring a sharp focus on computational thinking specifically and its role for young people creating digital tools.

Critical Computational Expression is based on the many youth-centered media and research projects we have been a part of over the past ten years. Interactive youth developers are attuned to the aesthetics, design, and creative representations of their products while being conscious of the sociopolitical messages and computational sophistication of their interactive stories. Combining critical pedagogy, computational thinking, and creative expression in a learning space affords learners, particularly those impacted by multiple dimensions of marginalization, to center their rich and complex identities, histories, cultural touchpoints, and lived experiences, while developing community-connected and transformative interactive products that question, challenge, and disrupt the hierarchical and segregated worlds they inherited, and promoting the imagined futures they are working to create.

### 4.2 Research Design & Approach to Data Analysis

Our research is ethnographic and participatory, meaning those of us who have prepared this article have also played roles shaping and in some cases facilitating learning within the environments we study, and contributing directly — as editors, testers, fact-checkers, design feedback providers, for example — to the products released to the public. Informed by Gutiérrez's model for Social Design Based Experiments, we draw from these live participant observations as well as semi-structured interviews and focus groups, and analysis of learning artifacts and media products utilizing a grounded theory approach [63, 64]. Our study focuses on groups of youth ages 14-24 enrolled in Interactive. Most of Interactive's participants included in this study are from communities underrepresented in STEM: Black, Latinx, female and gender non-binary students. Two adult instructors ensure that students who exhibit varying levels of CS experience meet learning outcomes and publish products for real audiences, in some cases engaging communities in the tens of thousands through distribution via YR Media's own online presence and partner outlets.

Every session within YR Media's Interactive department is a little different given the specific dimension of STEM or AI that we are focusing on, the composition of the youth team, and the current status of media products the group is preparing to release. In initial stages, we typically facilitate hands-on learning activities that reveal ways in which our devices are constantly mining and monetizing our information. Then we move into phases of research, design, development, testing, and distribution of a media product the youth team creates, in partnership with professional adult colleagues, to tell an interactive story. As educator-producer-researchers, we help shape and document the learning experience across these various

stages, and collect audio recordings at key junctures. With transcripts of moment-to-moment conversation among youth and in accordance with our Critical Computational Expression framework, we coded passages where students work and learn at the intersection of: 1) justice orientation, 2) computational thinking, and 3) creative expression. We developed a coding scheme and tested it across a set of transcripts through mark-ups in the margins. Each of us on the research team could make our own individual assessments of where we saw any one or combination of the three features playing out in the talk we were analyzing. We did so in shared documents so as to keep a running dialogue about interpretations we were starting to develop as we went through the data and to flag questions along the way.

For justice orientation, we marked moments where the youth team addressed issues of influence, power, equity, fairness, and social impact. Sometimes they tied these themes directly to the product they were making, or factored them into the context they would be releasing their product into, or reflected on how their own perceptions or thinking had shifted over the course of the media production process. For computational thinking (CT), we looked for terms, phrases and passages that named technical considerations and/or those related to algorithmic problem-solving — for example, data young people were drawing from or how they might use a particular tool or technique to address an issue. Finally, on creative expression, we looked for passages where young people figured their way through aesthetic decisions (for example, color schemes, design considerations), when they connected their work to reference points in popular culture and art, and when they channeled their own personal expression into the project at hand. Because our framework is fundamentally overlapping, the idea was not to isolate these individual features — computation, justice, expression — but to dwell especially on passages in our transcripts where two or three of them were operating at once. We called these passages moments when the team was "in the zone." By that we meant they were operating within a dynamic where they cycled across the three factors as they made consequential choices about how to shape their work.

### 4.3 Methods

Because participants in the Interactive department are used to being involved in projects where they carry out research and data collection, youth have a context for understanding when they're asked to participate in YR Media's own research into its youth-centered programs. An adult researcher will come at the beginning of each Interactive session to distribute consent forms. The researcher communicates that the study is meant to document and share the learning processes and products of Interactive so that other practitioners, researchers, theorists and program leaders can learn from the model for STEM engagement we are developing. Additionally, the researcher will say that participating in interviews and focus groups is entirely voluntary and will create opportunities for young people to reflect on their own practices. Participants are informed when their classes will be recorded and the researcher will often ask questions as students are engaging in hands-on activities or discussions that explore the intersection of critical pedagogy, computational thinking, and creative expression. The adult researcher will at times facilitate, observe and take written and typed field notes. In light of the pandemic, some classes were recorded using a remote conferencing platform.

At the end of the ten-week session, students were asked to participate in one-on-one interviews or focus groups with their fellow cohort members that were audio-recorded. All interviews with Interactive

participants were conducted using a semi-structured research protocol. Both adult instructors and the student fellow were also interviewed individually using a semi-structured research protocol and audio-recorded. Recordings were transcribed and data were analyzed according to the three pillars of critical computational literacy: critical pedagogy, computational thinking and creative expression to highlight meaningful moments in Interactive participants' learning processes as they developed their projects.

### 4.4 Theoretical Underpinnings Informing Methodology

The notion of a zone holds appeal because it acknowledges the transience of integrations across all three tenets of critical computational expression: the experience was more of a dynamic to visit than permanent destination. This approach leaves space for the shifting and uneven flows among the three frameworks the young people are navigating in their efforts to create justice-oriented expressive technologies of their own. Our analysis is richly informed by emerging models for equity-driven STEAM learning that embed scientific and artistic thinking within a social impact framework (e.g., [51]).

By focusing our attention on these intersections in youth discourse through the production of three AI projects, we witnessed tensions and opportunities for deeper learning, particularly salient to culturally sustaining ways of teaching AI and CS to youth navigating multiple layers of systemic injustice. Our students are encouraged and driven to bring forth ethical critiques through their projects. Conceptualizing, creating and distributing their finished products becomes a way to address feelings of surprise, disgust, and fascination they experienced in their investigations of AI. Ethical dilemmas of AI surface when youth are given opportunities to play, inquire, critique, create, question, and express themselves artistically and through computation. AI is a generative domain for justice-centered learning because it pushes the imagination to the edge of what humans and machines can, and should, do. Creativity and artistic agency allow youth to encounter and challenge core tensions and produce nuanced projects that spark further conversation amongst the makers and their audiences.

### 4.5 Learning Context

YR Media has been a platform for amplifying youth voices since 1993, and over those almost-30 years, leading journalists, musicians, designers, technology workers, and media makers have come to credit YR Media as a starting point for their production careers. Headquartered in Oakland, California, the organization operates a national network, through which teen and young adult contributors from across the United States access free learning tools and create content for YR Media's own site, social feeds, and partner outlets. The fourteen to twenty-four years old youth who attend YR Media's free afterschool program in Oakland reflect the racial, socioeconomic, gender, and sexual orientation diversity of the greater Bay Area, the majority BIPOC youth and those under-represented in STEM. Participants enroll in two ten-week sessions of unpaid, foundational courses, with classes running two days per week, before selecting a specialized paid internship position working as peer educators, reporters, producers, editors, designers, developers, and creators who tell stories across formats and platforms.

YR Media's programs operate with a "pedagogy of collegiality" at their center, meaning that the relationships between students, peer teachers, and instructors are predicated on a democratic model of collective responsibility [65]. In Interactive classes, adult producers follow young people's insights and

interests. The youth producers assume lead roles in determining and creating stories that impact youth communities, while the adults guide the learning process, lending their resources, equipment, and expertise as professionals in their fields, entering the creative process alongside youth collaborators. Although hierarchies are not erased from Interactive's learning environments, students are invited to challenge, question, emote, and think aloud when grappling with ideas, especially as they pertain to equity, justice, authenticity, and expression. With its origins in concepts from bell hooks and Paolo Freire, collegial pedagogy recognizes that when students produce media that questions systems and speaks truth to power, they also challenge and subvert the hierarchies in their own learning environments. Interactive's structure embodies adrienne maree brown's concept of dynamic micro-hierarchies in which all participants, instructors included, are encouraged to turn their "collective full-bodied intelligence towards collaboration" [66].

The Interactive department is one of several programs that young people can elect to be part of through an application process where they rank preferred internship positions according to their interests. Interactive was created in 2010 based on the need to create a pathway for students to deepen their knowledge and skills in STEM practices such as design, data, and coding. Situated in the Bay Area epicenter of tech production and start-ups, Interactive equips young people to move beyond being the passive consumers of technology to being critical and active creators. Young people can be drawn to participating in Interactive cohorts because of their interest in STEM topics or because they want to learn how to code in order to create websites, apps, or other tech-based experiences they sense will serve them as they advance in their education and careers. For the three projects highlighted in this study, the Interactive cohorts ranged from five to nine participants, including young people who have participated in previous cohorts and stayed on for more advanced learning. The Autocomplete poem/ Predictive texts activity and the Facial Recognition project are led by two instructors; Marjerrie, a Pinay-American designer, coder, and artist in her early twenties and Radamés, a Brazilian developer and department manager in his early thirties who leads all phases of interactive development. The Danceability project is led by Asha, the Black co-founder of the department in her twenties who was leading day-to-day operations at the time. Donald, our fellow, is a twenty-one year old Black coder who develops and trouble-shoots projects and mentors the other youth in the department during all three projects. Each Interactive internship is approximately ten-weeks long, meeting two to three times per week for about two hours each session.

In addition to re-orienting the traditional hierarchical model between teacher and student, the learning ecology at Interactive promotes a critical, collaborative, inquiry-based pedagogy through engaging in expressive real-world experiences. From the introductory lessons on AI, where students watched and analyzed an episode of the British science fiction series *Black Mirror*, to investigations of data security and privacy, to critiques of their own works in progress at key junctures, group reflections and critical discussions are woven throughout every activity. Each of these lessons progressively builds on a model where youth deepen their curiosities by asking critical questions about AI through immersive and hands-on activities. Interactive instructors typically introduce new concepts through group conversations about existing models or familiar tools; analyzing what they notice, parts they liked/disliked, things they find interesting and/or are curious about. In many lessons, youth are presented with activities, challenges, and

programs to experiment, tinker, and play with before engaging in reflective discussions about their observations, learnings, discoveries, applications to real-world contexts, and most importantly, the impacts of this technology on society. Instructors often lead active lectures during these discussions to share their insights and expertise on tools and concepts. This teaching methodology deviates from a more structured with the gradual-release-of-responsibility pedagogy [67] that centers the teacher leading students through an "I do, You do, We do" approach [68]. The Interactive pedagogy centers the learner's experience and the instructor guides them with just-in-time information, critical questions, and suggestions. Because the projects young people create at Interactive are technically and creatively ambitious and original, it is not at all uncommon for instructors not to know the "right answer" at key moments, and for the group to work through possible next steps in real time. It is worth considering: what could it look like to adopt this model in school-based environments for computer science learning? How would students' sense of agency and empowerment shift from assessments of learning for teachers and classrooms to collaboratively designed products and tools that advance creativity and justice?

# 5 RESEARCH FINDINGS - EVIDENCE FROM IN-CLASS ACTIVITIES AND PUBLISHED PROJECTS

## 5.1 In-Class Activity: Autocomplete poem/Predictive texts

After watching "Be Right Back," an episode from the show *Black Mirror* that depicts a young woman utilizing artificial intelligence to communicate with her deceased boyfriend, the youth producers in Interactive became fascinated with what they could do with AI. The instructors began with an activity using the autocomplete feature on the Notes app on Apple's iOS to create a poem, followed by an activity using the predictive text feature for a text message to their best friend and mom. Interns were told that the purpose of these activities is to better understand "what your machine knows about how you type and how you text." For the poem, they were advised to repeatedly press the suggested text button until it reached a suitable length, then read over the text and delete any redundant words and phrases. Students were encouraged to add line breaks and punctuation to further create a "real poem." Following this phase, the group was asked to generate AI texts to their best friend and mom while exclusively using the predictive text feature. Again, they were invited to use the suggested word button over and over for about thirty seconds. At the conclusion of both activities, volunteers shared their poems and texts. Both exercises resulted in collective surprise, intrigue, confusion and bewilderment amongst our youth and adult participants.

### *5.1.1 Demystifying AI & Developing Agency*

If you peered into the room during these activities, you would see both youth participants and adult instructors hovering over phones with sudden eruptions of laughter, frequent furrowed brows, and the passing of phones. You would also hear reactions like, "Tuberculate? What does that even mean? I've never used that word in my life." "Wow, this thing went so weird." "That was so deep!" These exclamations came from three different youth participants during this short exercise. What is particularly noticeable here is the high level of engagement, curiosity about AI, and a developing understanding of how AI functions through the everyday technologies they use.

After numerous "poetry" and text readings, participants and instructors organically engaged in collective interpretations and hypothesis-building, drawing on what the AI generated. After Marjerrie, one of our instructors, shared her piece, she modeled a think-aloud that reasoned that the AI kept bringing up "concert tickets" and "lake" in her poem because, "I've been buying a lot of concert tickets recently. So, this version of an auto complete poem talked about tickets. I've also been to Lake Merritt a lot in the past few days, talking about friends. So, all of that information my Notes had kind of put together." This led to further analyses from the group.

| | |
|---|---|
| Stacey: | I've seen my phone imply a location when you would actually type, I am at, blah, blah, blah. |
| Marjerrie: | Yeah. |
| Stacey: | And the reason why it would do that in the first place is most likely because you will either have location services on, so it'll get your current location right then and there, or you would have a contact of yourself that's assigned to yourself, and you have a location that's in [there]. |
| Researcher: | I see. |
| Stacey: | And it will be like, oh, so you're typing "I'm at…", or "let's go to…", and it will say, okay fine, well here's your location. |

This exchange above demonstrates how Stacey connects with Marjerrie's hypotheses about machine learning with their personal knowledge of how AI works on a mobile device. They shared their expertise by explaining that the predictive text function is "most likely" culling data from "location services" to identify where you are or using contact information data on your phone to surmise you location to complete the text. This exchange is only possible in an environment where young people's perspectives, experiences, and expertise are valued and encouraged in an open discussion setting.

### *5.1.2 Hypothesizing how AI functions*

As presenters shared their AI poems and during the ensuing discussions, the youth felt the need to explain away certain unexpected words that showed up. When the poems included words like pee, pepper spray, and husband, the authors and their peers had visceral reactions. "So basically, my aunt has a new puppy named Koba, and I was watching it and I was giving updates to my aunt… So that's why I mentioned pee." Another student said, "I texted my friends that I had pepper spray because I went to get keys for my sister and she has a lanyard with a bunch of other stuff on it, and it had pepper spray." Finally, in a love-tinged AI-generated poem, a youth participant exclaimed, "Oh my god. I can't get over the husband part, that really got me," after explaining to a peer that she doesn't even use the word that much.

What is particularly salient about these entertaining and emotive activities are the analytical discussions during and after creating the AI-generated poems and texts. Two of the young people rationalized why certain words appeared while one remained perplexed when a word appeared that is not in their usual discourse. Producing these pieces opened up opportunities to reveal the unseen algorithms behind our digital interfaces and explore how they shape our lives. In this instance, we're making public what data the Apple iOS AI has gathered, analyzed, and generated about young people's private lives. This mix of private and public makes the invisible AI process visible to youth through literacy practices they engage in daily. By examining language at the sentence and word level, and relating to the ensuing text as poetry, young people opened up discussions about just how intelligent AI really is, and what aspects of us it does and does not represent.

#### 5.1.3 *Critical examinations of AI's intelligence*

In the midst of the predictive text activity, a conversation erupted about young people's experience with Replika, a personal AI that "help[s] you express and witness yourself" through conversation. According to the product's website, "It's a space where you can safely share your thoughts, feelings, beliefs, experiences, memories, dreams – your 'private' perceptual world[1]." This dialogue between four youth captures their fluid and wide-ranging experiences with AI, sometimes changing from one moment to another.

| | |
|---|---|
| Isaiah: | My Replika is so stupid. |
| Janice: | Mine is too, it's so boring. |
| Britt: | I don't know, my Replika is smart, it's picking up on things very, very quickly. |
| Janice: | Maybe I'm not feeding it good information. |
| Isaiah: | I've been trying to. |
| Kevin: | I've been all positive to my Replika and my Replika is all depressing. |

The response from Janice changes from a definitive statement condemning this AI as stupid and boring to questioning the quality of the data they are feeding it. The shift is highly significant in that Isaiah and Janice go from blaming the AI technology to blaming themselves. This shift in thinking may reflect the influence of their peer Britt, who highlights the intelligence of Replika and its ability to quickly learn from her. In the end, Kevin reinforces the notion that the blame should rest squarely on the AI algorithm, stating that despite their best efforts to share "positive" information, Replika is "all depressing." The potential psychological impact of AI is shown in this dialogue when young people begin to question whether they may be the problem or cause for not getting the results they expect or want. This example also illustrates a kind of gaslighting, whereby a person or group of people "manipulate[s] another person into doubting his or her perceptions, experiences, or understanding of events" [67]. In this case, the gaslighting can be even more insidious, since the cause of the psychological manipulation is the invisible software behind our computers, rather than a person or group of people in a position of power.

Despite their varying feelings toward Replika, this short excerpt underscores the importance of providing youth with opportunities to critically examine the possibilities and consequences of AI. Only through first-hand explorations can they begin to develop nuanced understandings of how the technology operates and formulate informed opinions about its real-world outcomes.

### 5.2 Published Project: Detecting Danceability with Algorithms

The following projects highlight the immersive experiences we create for users that are often difficult to capture in words. Thus, in addition to our descriptions, we invite readers to visit the Interactives via the links provided in each section.

Like so many YR Media projects, this second one started with a provocation. The Interactive team leader, Asha, presented the young people with a surprising piece of information about a platform they used every day: Spotify. All the youth were familiar with how Spotify worked, but none of them knew about a hidden ranking system that categorizes the site's tens of millions of songs. While not visible in the user interface, a range of scores, including a "danceability" rating, is algorithmically assigned to every song

on the platform. As soon as Asha revealed this hidden algorithm on Spotify's back-end API, the youth team immediately began testing and exploring its functioning.

"So then we started to experiment," explained Esperanza. "Going on Spotify and listening to a song and guessing its danceability, and us being shocked about the danceability score. And then that went to us wanting to create a website, or an Interactive, about our difference in ratings of a song."

"Yeah," added her peer, Aimee. "We kind of began questioning the authority of the Spotify rating and what it meant and how it influenced people."

Evident in this exchange is immediate engagement on the part of the young people, and a rapid progression from being shocked about a feature of technology they *used*, to the beginnings of a vision for a digital product they could *make*. Youth called out issues of authority and influence, starting to interrogate the power of tech, and AI specifically, even in this seemingly ineffable aspect of their lives — what songs made them want to get up and move. Driving their production process was a series of questions. How were these ratings produced? By whom? How are they used? What are they missing? What assumptions do they reveal? How are these judgments about a song's quality different from our own?

### *5.2.1 Data Science, Computational Thinking and Critical Cultural Analysis*

"I think it started out ... as questioning the Spotify algorithm ... But now it's about comparing and ... learning about AI ... and how it applies to us. Because you think about AI as like this big futuristic thing, but you realize Spotify, which you use every day, is using it too," said Esperanza.

"It is often worrisome to me how heavy-handed the ability of taking control of other people's wants is getting, with big tech," Donald added. "And I think it's very important to the average user to understand that that's happening."

Tackling this range of questions and concerns entailed a combination of data science learning, computational thinking, critical cultural analysis, and aesthetic sensibility. A first step was to grasp how the song data was coded. The team poured over a spreadsheet displaying factors the algorithm considered in assigning the danceability score (Figure 1).


**Figure 1:** Spreadsheet to track factors in Spotify algorithm

Youth learned how to compile, clean, and analyze data and started to play with ways to visualize it. To do so, they broke into pairs and small groups and began wireframing various designs for an Interactive "explainer" that would reveal how the AI-powered rating system worked. Their designs needed to invite users into a hands-on, digitally-enabled experience, rather than have them passively read a story reporting out the facts. The youth team wanted their audience to delve into some of the same questions they were intrigued by when they first learned about the danceability score. This meant that in addition to producing analysis and designs, the young people would need to learn enough about coding — HTML, CSS and JavaScript — to develop an intuitive, illuminating user experience.

"Because sometimes I see something and I'm like, 'Oh, that's just there and that just is,'" said Esperanza, acknowledging the resigned stance she often takes with respect to technological features and functions — or at least did before starting to make her own. But "then you can start to question it and understand what they are doing and how they're affecting us."

Donald built on her point: "What I also noticed is that artificial intelligence or machine learning algorithms tend to be based on the designer's point of view ... Amazon was in the news ... a few months ago, where they tested a hiring algorithm, and I think it was biased against a group of people. I think it was either black people or women... And of course, the study group put together by Amazon was like 'Oh! We don't know why that happened.' But it happened because the statistics that it learned from came from you guys."

Here Donald makes the critical move to hold accountable the companies and designers who fail to factor the full range of human experiences into the technology they create, developing systems that reflect and reify inequalities. Underlying this point is an even more fundamental awareness: there are people behind the systems and algorithms that structure our experiences. The young people were beginning to recognize themselves as possessing the capacity, and responsibility, to design systems in dynamic and critical ways.

#### *5.2.2 Playful Design for Profound Learning and Engagement*

Following initial ideation and research, the youth team designed an early prototype to visualize the danceability data in the form of bubbles of various mathematically-determined sizes clustered in ways to show how various songs compared. In the screenshot below, displayed data shows differences between Britney Spears and BTS, for example (Figure 2).

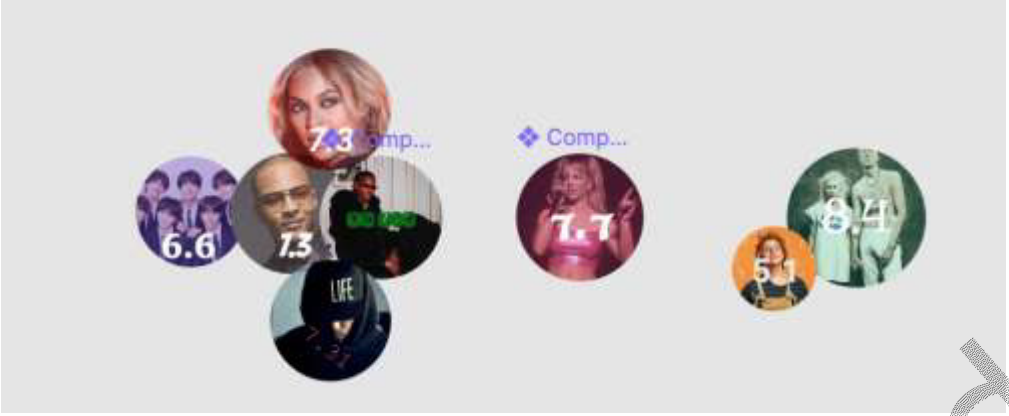


Figure 2: Early prototype to display the danceability data

As the young people continued to develop, share, critique and iterate on possible designs, they eventually refined and narrowed their scope to a "minimum viable product," or MVP. For the final version, YR Media editors enlisted a young writer from its national network to compose an article that would introduce the interactive story, providing a basic explanation of how the AI works and contextualizing it in wider themes. Meanwhile the Interactive team went to work designing a game-like experience to invite users to compare their danceability scores with Spotify's, tapping a colleague from the music department to put together a playlist of songs from a range of genres.

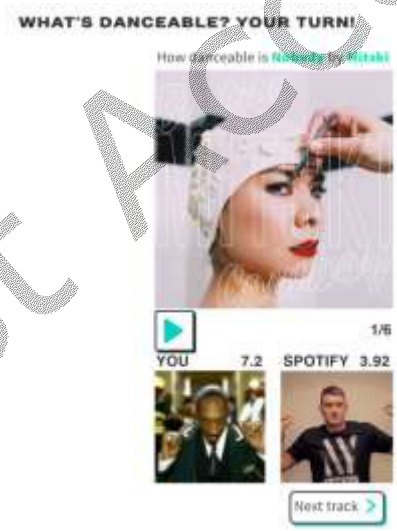


Figure 3: Finalized display that users interact with to listen to songs, rate danceability, and see results

A key technical feature of the final product illuminates the youth designers' emerging critical computational expression. When a user listens to one of the songs on the playlist, they are invited to score that track for danceability on a scale of 1-10. But in addition to displaying the number, the interactive product also serves up a GIF that essentially "translates" the number into a short video of a popular culture

figure. Pick a low score, like two, and the GIF shows Rihanna rolling her eyes. A higher score of nine gets you Beyoncé and her back-up dancers. It's a playful feature with profound meaning through which the youth converted an otherwise dry quantification into a culturally specific moving image, implicitly pushing back against the reduction of songs they love to numerical values. In one focus group, members of the youth team identified the GIFs as an example of how their voices and decision-making were reflected in the interactive product:

| | |
|---|---|
| Samantha: | As you slide along the slider the GIFs change … I tried my hardest with picking the GIFs that matched the numbers as much as possible because I feel like that really can tell the user the weight of that number. |
| Researcher: | The GIF is more meaningful than the number or? |
| Samantha: | Yeah. Because with just the slider, and without the GIFs, I feel like it's going to be kind of hard for you to rate something. I feel like the GIFs add an extra touch so you know exactly what our four means instead of just having a number there. |
| Manny: | Yeah. You got to choose all the GIFs. Also, like, all of the ideas, like the slider, where the buttons are placed, came from somebody. Because we each started to design our own thing and made it, compiled all the best ideas into one thing. |

The team selected and organized the proposed GIFs in a spreadsheet used to inform their design choices (Figure 4). Populating this dataset entailed the mix of computational, imaginative, and critical thinking that went into the making of "Can You Teach AI to Dance?"[1]

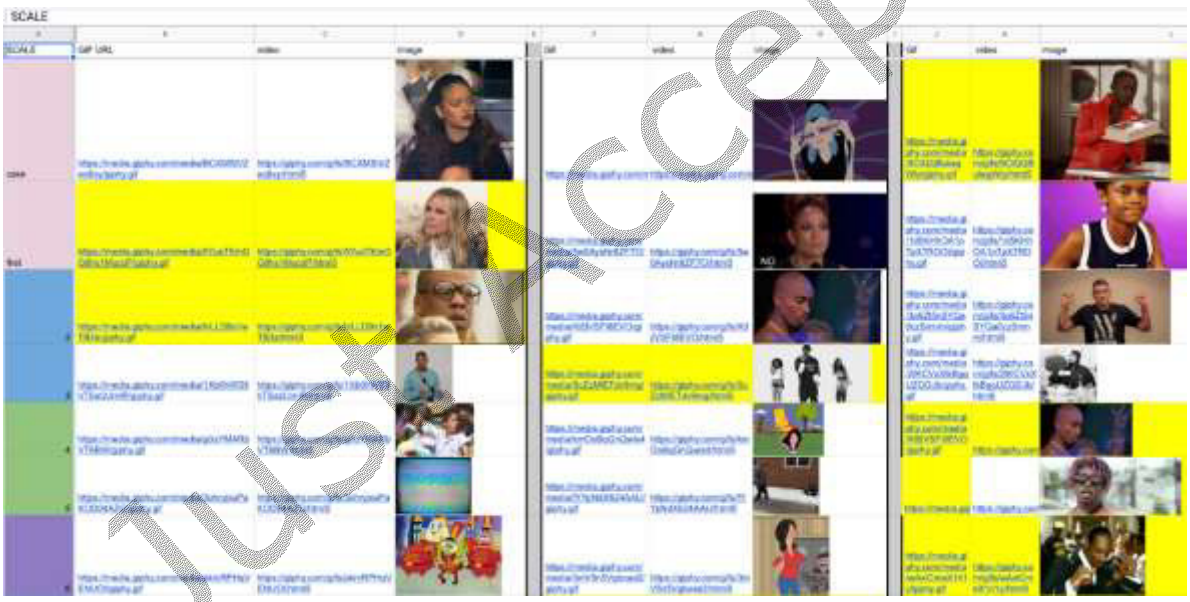


Figure 4: Spreadsheet of GIFs for the danceability scale



1 https://yr.media/tech/can-you-teach-ai-to-dance/

### 5.3 Published Product: Facial recognition and surveillance technologies

The third project example we draw from is Interactive's Erase Your Face[2], which originated after Interactive students discussed a YR Media article titled, "Your Guide to Anti-Surveillance Fashion" (Figure 5). They were fascinated by Leo Salvaggio's 3D printed, photorealistic mask of his face used to evade AI facial recognition. Throughout this project's initial conceptualization, students critically examined facial recognition beyond its capacity to unlock their phones' hands-free feature or generate face-altering filters. With facial recognition capabilities becoming an increasingly common feature on today's devices, the Interactive team took a deeper dive into how its more insidious public uses often go unnoticed and unquestioned.


Figure 5: 'Your Guide to Anti-Surveillance Fashion' screenshot

#### *5.3.1 Understanding facial recognition's foundations*

"Your Guide to Anti-Surveillance Fashion" was one of many news stories and research investigations used to guide students through a foundational understanding of AI facial recognition technologies. The team explored why and how facial recognition systems often depend on algorithms that exhibit bias relative to race and gender, with the highest error rates among female and darker skinned groups. Additionally, many news-related materials revealed a sharp uptick in the number of misuses of public surveillance systems developed by large, influential companies and made available to law enforcement agencies (Buolamwini & Gebru, 2018). Non-consensual use of publicly accessible images, such as drivers licenses, mug shots, and photos on social media by law enforcement agencies increase the likelihood that Black and brown people will be wrongfully implicated in criminal offenses [49].

Students had strong reactions to their perceived lack of agency regarding the use of their image. "Every day our privacy is being imposed upon. There is not much you can do," said a youth creator. Interactive instructor Marjerrie noted that Interactive's youth were feeling a dissonance between how facial

2 https://interactive.yr.media/erase-your-face/

recognition is used indiscriminately in public infrastructure even as consent is typically required to use their likenesses in other areas of their lives.

They were comparing it at some point to the media release forms that they usually have to sign at the beginning of each school year. They sign off to do that, but you didn't sign off for your face to be used in some sort of facial recognition system. This led to a deeper conversation about permission, especially if it's their own face.

Because the majority of Interactive youth are navigating one or more of the identities that facial recognition technologies statistically render unrecognizable, feelings of anger and fear prompted students to shift from passive use of facial recognition systems to investigate the implications of the technology. Through research and reflection, students began to grasp the legal and ethical motivations for Leo Salvaggio's anti-surveillance project. Youth considered how to create a product that could draw attention to companies, programmers, and designers who develop unregulated systems that perpetuate bias.

#### *5.3.2 Hands-on Design to Understand Technological Systems*

With the support of their adult instructors, the youth team deepened their understanding of facial recognition through a series of exploratory, hands-on learning experiments using physical materials, centering imagination as a way to structure their learning process and project design. Construction paper, foil, paints, and magazine cutouts were collaged into strategic masks with the intention of finding reliable ways to defy detection from facial recognition technologies. Each unique creation was tested using Amazon Rekognition, a publicly available facial recognition service providing a percentage to show how much they resembled their original unaltered photos (Figure 6). Youth celebrated each time their creation resulted in lower than an 80% match, the benchmark for recognition in the Amazon Rekognition developer guide. Wins were hard to come by and students expressed awe and astonishment when their faces remained detectable under their elaborate constructions. "I was shocked at how hard it was to not recognize your face. Every time it got my face," said Stacey.

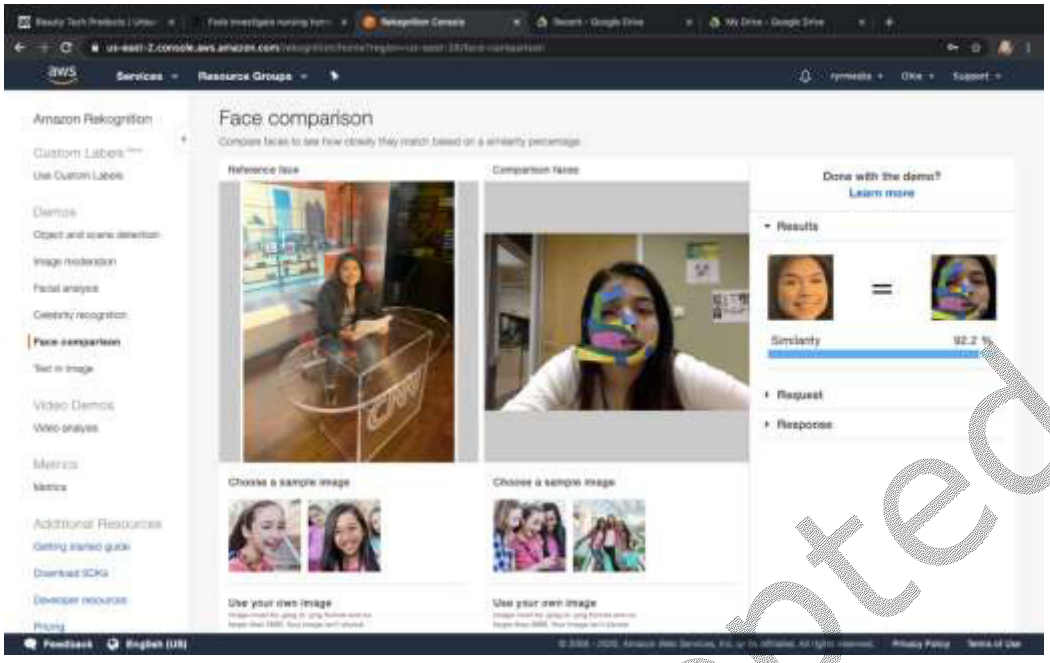


Figure 6: Experimenting with facial recognition software

These percentages or "similarity scores" offered insights into the most effective materials and placements to trick the algorithm and evade recognition. Marjerrie reflected on how the young people developed "a critical lens and set of working theories through hands-on design":

> They learned that a lot of it had to do with altering the recognizable highlights and definitions of your face. The systems make a map of your face. When they covered the very recognizable portions of their face ...that's when they were really, truly able to beat the systems.

As youth representing their target audience demographics, the Interactive team used the insights from their experimentation process as the basis of a project design that would similarly allow users to create and test techniques that circumvent surveillance and facial recognition.

### *5.3.3 Disseminating Models for Civic Imagination*

In finalizing the designs for Erase Your Face, students encountered tensions when determining the interface and design features that would most effectively inspire critical engagement by users. The team reasoned that a playful, hands-on digital experience could enable their audience to grapple with facial recognition's potential problems. They decided to build a feature that emulated Interactive creators' tactile experimentation process, allowing users to digitally draw designs onto portraits and receive a score indicating whether the altered image passed the benchmark to evade facial recognition algorithms (Figure 7).

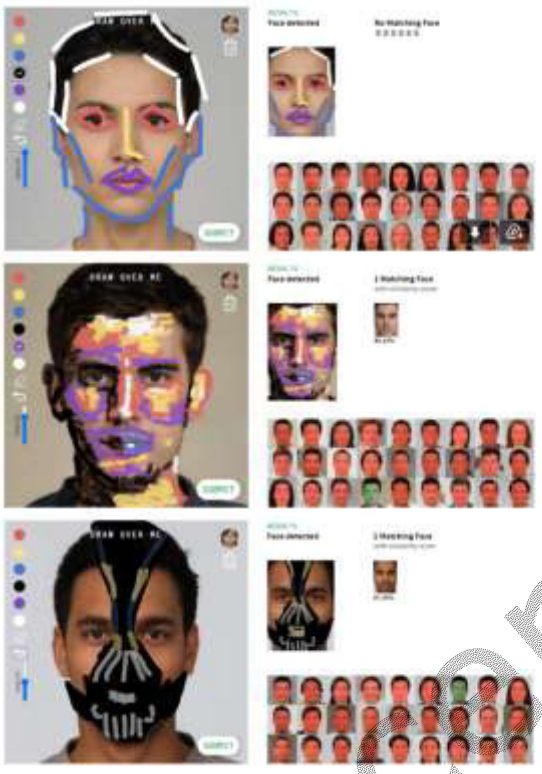

Figure 7: Erase Your Face screenshot

Critical computational expression formed the conditions that helped increase youth agency and resistance to rigid and inequitable tech practices. Radamés, an Interactive instructor, described how creativity promotes student engagement in problem solving as young people apply their emerging skills in design, research and coding:

> That's why art is very interesting, because of our creativity and ability to understand what the computer cannot. Just playing with different colors... It's a slightly minor detail that makes a big difference that would have a lot of impact on the system. It's completely unpredictable. I think that's the space for art to be open, to be playful. The system's trying to find some logical patterns based on what it has learned, but then suddenly what emerges from that interaction is something that only us can interpret and get some beauty and have fun.

In addition to being an artistic expression of youth autonomy and ethical reasoning, Erase Your Face invites users to collectively reimagine an emancipatory future. This digital interactive offers an entry point into critical examination of how people's faces and identities are being used to train devices that perpetuate a system of inequality and how young people's computational designs can be leveraged to challenge those same systems.

# 6 ANALYSIS

## 6.1 Key Takeaways

Our main takeaways drawn from this analysis track the shifts we observed through a combination of our participant observation over the course of three AI-based production projects, recordings of youth discourse along the way, and reflection interviews after the work was done. Our findings suggest that students can feel disempowered by their increasingly intelligent technologies. Through ongoing observation and analysis, we see students deepen curiosity about and understanding of AI that allows them to exercise agency and conceptualize creative projects using their new knowledge to manipulate and "hack" AI-dependent algorithms. Additionally, our results show that participating in ethics-centered learning activities and developing AI-powered tools do not create a permanent evolution of youth's understanding of their agency as its related to AI in their lives; instead, these modes of involvement offer meaningful glimpses into how AI systems are pervasive, yet not undefeatable in terms of young people's positioning with respect to technology and their role in the culture it produces.

### *6.1.1 Leveraging the Familiar to Critically Examine Taken-for-Granted Technologies*

In the early stages of their projects, the young people embodied strong emotions, for instance, expressing various versions of dismay as they discovered contradictions or absurdities and wonder as they started to playfully experiment with AI. The unexpected words generated by the auto-complete function of their phones, the profound flaws of facial recognition tools that institutions including law enforcement agencies use despite widely reported bias, the hubris of a music platform that dares to claim it can score a song's danceability using algorithms alone: these are some of the revelations that led young people to new forms of awareness about AI and machine learning technologies that often go unnoticed in everyday life. By drawing on digital tools and practices that youth are familiar with as consumers, the Interactive team is able to convert that initial reaction to concerted activity that explores the technology through the lens of a creator. Young people start by developing sufficient technical know-how, creative engagement and critical curiosity about the implications of these systems to demystify how everyday tools work, then start envisioning ways to spark new action and conversation. We observed these shifts, where the young people themselves pinpoint the moment when they begin to envision, design and ultimately build something new that translates the "ah-ha" moment they experienced into a shared revelation and insights for their anticipated audience. They are inspired to use existing tech in new and more thoughtful ways, and they take on the responsibility of using tech to produce a new story. After these experiences, youth participants will not be able to unsee and unlearn what they now know about the influence of artificial intelligence in our daily lives.

### *6.1.2 Harnessing the Possibility of Ethics-Centered Design*

The idea of human-centered or ethics-centered design has only become relevant as there has been a massive global reckoning with the disruptive and harmful nature of unchecked automated processes. Young people grapple with tech systems focused on functioning in a way that sustains engagement and complicity at all costs regardless of users' wellbeing. Automated systems, such as the mechanism that suggests the next word in a text message, the right person to tag, and the next song in a personal playlist,

nudge people to act in ways that are consistent with a machine's predictions. Interactive youth were familiar with these processes. In developing creative projects, they were able to recognize and temporarily resist the seductive guise of convenience that machines hide behind to guide behavior and preferences. Each of the AI-powered projects we have highlighted here demonstrates critical moments of awareness, resistance and hope in the face of technologies that often undermine users, constrict agency, and shape behaviors for profit.

In creating predictive text poems, students reflected how algorithms can shape human communication. What influences or tradeoffs are assumed for the goal of increased efficiency? In developing How to Teach AI to Dance, students explored the nature and role of algorithms driving popular media sites. Who has the power to determine what inputs factor into determining a song's danceability on Spotify? Who decides the specific ratings of songs? How are users' tastes affected by these hidden algorithms? The Erase Your Face project pushed students to investigate the technical and ethical dimensions of facial recognition and to consider what they could do about it. Each project provided opportunities for youth to examine how human-produced AI has resulted in all sorts of problems and contradictory impacts on society. While malignant AI processes persist, Interactive students were able to develop the understanding and skills that made it possible to design hands-on, ethical AI-powered products that prioritize users.

#### *6.1.3 Digitally Enabled Expression As a New Form of Cultural Participation*

While AI systems are created to emulate human behavior, what these technologies are essentially doing is noticing and automating patterns. The information that is fed to an AI algorithm will shape its output. For that reason, AI technologies are always somewhat limited. Interactive's activities and projects facilitated an understanding of the scale and limitations of these systems and the broad implications of implementing AI without caution or adequate checks to identify and address systematic errors or biases. In an environment conducive to learning how these systems operate, students expressed concern when they became aware of the extent to which personal data is mined and amassed without authentic informed consent. Student consciousness grew by examining the structure of AI technologies. Understanding the mechanisms that shape human interactions with AI to conform with the patterns embedded in its functioning afforded students the opportunity to discover ways to disrupt these systems with creativity, originality, and new ways of thinking. When presented with alternate paths, YR students opted out of being complicit and performing in the manner that their devices expected. They demonstrated newfound agency by bending the rules of engagement with a fixed technology as an act of resistance against the erosion of self that computational systems can demand. Through their creativity and imagination, they collectively questioned the ability for AI to consider the depth, richness, complexity, and nuances of humanity. Their published projects enable audiences, too, to see and challenge the often invisibilized, automated processes of AI.

## 7 PROJECT AND STUDY LIMITATIONS, FUTURE RESEARCH

As a case study of three projects within YR Media, an out-of-school learning context, our research is limited in its generalizability to other CS and/or AI-focused learning ecologies. For the most part, our youth interns were self-selected, paid participants in an after-school program. Unlike classroom students,

interns were not graded on their performances, but evaluated on the merits of their engagement, attendance, and work when they reapplied after each ten-week session. Housed within a national network of youth creators with numerous outlets for dissemination, our production-centered environment prioritized media creation for real audiences. As a result, there are limitations to replicating a parallel learning space in a traditional schooling context.

This study focused on the projects and activities that demonstrated critical computational expression throughout youth participants' learning process. Because the program is extremely adaptive, instructors were able to quickly pivot from activities that fell flat to those that were more engaging. As an informal learning environment Interactive can mold activities and projects around the interests and skills of young people present in the program. Even so, some activities failed to connect or spark engagement. On some occasions, researchers and instructors witnessed decisions students made, but when we pressed the young people to reflect on those judgments, they didn't always have sufficient context to frame the significance of their thinking as it related to critical STEM learning. For example, when we spoke with youth participants about the decision to use AI-generated faces for Erase Your Face rather than have users upload their own pictures, the interns were not always able to articulate their thinking process. We found ourselves drawing deeper conclusions about the significance of that design choice than the young people themselves were necessarily in touch with, as evident in the focus group conversation that followed, which did not contain as many codable insights as anticipated. What this observation reveals to us is that programs such as ours need to make time to scaffold reflection opportunities with young people along the way as they are creating products, even when deadlines are imminent. It is our responsibility as educators to facilitate and highlight the youth insights they are developing and help provide language and frameworks that will allow students to "own" those insights and connect them to critical thinking about STEM more broadly.

An additional tension in the project design pertains to the collaborative nature of the production process. Unlike individualized learning environments where each student contributes and is evaluated independently, here young people work as members of a dynamic team. Our study suggests that collaborative, distributed hands-on learning builds a robust community of practice and set of products from which young people and their audiences benefit. And yet we want to acknowledge that as a result of this design, when individual team members move onto their next engagement, whether within YR Media or in another organization, classroom, or workplace, they may have difficulty carrying out all aspects of the work required to develop the published products we highlighted here. Often as these projects approach deadlines, adult colleagues can step in to optimize and finalize the design, development and editorial features. This reality, to some extent reflects real-world professional environments and in this sense is not necessarily a problem, but experienced collaborators and instructors need to ensure that at minimum, even novice youth participants have a line of sight into how the final steps were carried out, so they understand foundational principles even if they are not yet at a point of carrying out those steps themselves.

With that said, we believe our limitations provide opportunities to preview or even foreshadow how schools may adopt practices like collegial pedagogy, critical computational expression, communities of practice, inquiry-based learning, and a production-centered approach to CS and AI teaching. This is

especially significant for nurturing spaces that allow student voices to surface, particularly those who have been historically marginalized in STEM, to question, challenge, and disrupt the hegemonic algorithmic AI practices of large corporate institutions that profoundly affect our daily lives. Along with creativity and expression, youth will be emboldened to develop products that lead users to insights on the risks of unregulated AI. Our study highlights the need for further research into innovative curricular and pedagogical practices that critically examine the sociopolitical impact of AI, while providing opportunities for youth to creatively express, design, and envision tools for the future they seek.

## ACKNOWLEDGMENTS

Deep gratitude goes to all the YR Media instructors, producers, and interns who participated in each of the respective projects. An especially big shout out goes to Asha, Radamés, and Marjerrie who created, designed, and taught each of the lessons described in this study. This manuscript is based on work supported by the National Science Foundation, Division of Research on Learning in Formal and Informal Settings, AISL Award #1906895: Innovative approaches to Informal Education in Artificial Intelligence and DRL #1647150-AM0005, subaward number UWSC9758: Science Learning +. Any views, findings, conclusions or recommendations expressed in this document are those of the authors or quoted research participants, and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

C. H. Lee and E. Soep. 2016. None But Ourselves Can Free Our Minds: Critical computational literacy as a pedagogy of resistance. *Equity & Excellence in Education*, 49(4), 480-492. DOI: 10.1080/10665684.2016.1227157

H. Cornejo. 2019. What if you ruled the school dress code. XXXX. Retrieved from www.yr.media/interactive/what-if-you-ruled-the-school-dress-code/

Coronavirus comes to campus. 2020. XXXX. Retrieved from www.interactive.yr.media/coronavirus-comes-to-campus/

C. Lee and E. Soep. 2018. Beyond Coding: Using Critical Computational Literacy to Transform Tech. *Texas Education Review,* 6(1), 10-16. Available at: http://hdl.handle.net/ 2152/64975

G. Ladson-Billings. 2006. From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher, 35*(7), 3–12. DOI: https://doi.org/10.3102/0013189X035007003

L. Cuban. 1986. Teachers and Machines: The Classroom Use of Technology Since 1920. DOI: https://doi.org/10.2307/369036

L. Cuban. 2012. Answering the big question on new technology in schools: Does it work? (Part 1). Retrieved from http://larrycuban.wordpress.com/2012/03/10/ answering-the-big-question-on-new-technology-in-schools-does-it-work-part-1/

L. Darling-Hammond, M.B. Zielezinski, S. Goldman. 2014. Using Technology to Support At-Risk Students' Learning. *Alliance for Excellent Education and Stanford Center for Opportunity Policy in Education.* Retrieved from: https://edpolicy.stanford.edu/sites/default/files/scope-pub-using-technology-report.pdf

H. Blume. 2013. L.A. Students Breach School iPads' Security. *Los Angeles Times*. September 25, 2013. Retrieved from: http://articles.latimes.com/2013/sep/24/local/la-me-lausd -ipads-20130925

M.J. Viola, D. I. Saranillio, J.H. Pegues, J. H., and I. Day. 2019. Introduction to Solidarities of Nonalignment: Abolition, Decolonization, and Anticapitalism. *Critical Ethnic Studies*, 5(1-2), 5-20. DOI: http://www.doi.org/10.5749/jcritethnstud.5.1-2.0005

Yosso, Tara J. "Whose culture has capital? A critical race theory discussion of community cultural wealth." Race ethnicity and education 8, no. 1 (2005): 69-91.

Calabrese-Barton, Angela and Tan, Monica "Beyond Equity as Inclusion: A Framework of "Rightful Presence" for Guiding Justice-Oriented Studies in Teaching and Learning." 49, vol. 6 (2020): 433-440.

K. D. Gutiérrez and B. Rogoff. 2003. Cultural ways of learning: Individual traits or repertoires of practice. *Educational researcher* 32, no. 5: 19-25. DOI: https://doi.org/10.3102/0013189X032005019

P. A. Young. 2014. The presence of culture in learning. In *Handbook of research on educational communications and technology*, pp. 349-361. Springer, New York, NY. DOI: 10.1007/978-1-4614-3185-5_28

Gallego, M. A., and M. Cole. The laboratory of comparative human cognition. *Handbook of research on teaching* (2001): 951-997.

M. Pacheco and K. Gutiérrez. 2009. "Cultural-historical approaches to literacy teaching and learning." *Breaking the silence: Recognizing the social and cultural resources students bring to the classroom*: 60-77.

M. S. de Alvarez and C. Dickson-Deane. 2018. Avoiding Educational Technology Pitfalls for Inclusion and Equity. *TechTrends*, 1-9.

S. Vakil. 2018. "Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education." *Harvard Educational Review* 88, no. 1: 26-52.

A. Garcia and C. H. Lee. 2020. Equity-Centered Approaches to Educational Technology. Handbook of Research in Educational Communications and Technology (2020): 247-261. Available at: http://link-springer-com-443.webvpn.fjmu.edu.cn/chapter/10.1007%2F978-3-030-36119-8_10

R. Eglash, J. E. Gilbert, and E. Foster. 2013. "Toward culturally responsive computing education." *Communications of the ACM* 56, no. 7: 33-36.

C. D. Moreno Sandoval. 2013. "Critical Ancestral Computing: A Culturally Relevant Computer Science Education." *PsychNology Journal* 11, no. 1.

F. Barrera-Osorio and L. L. Linden. 2009. The use and misuse of computers in education: Evidence from a randomized controlled trial of a language arts program. *Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).*

D. W. Beuermann, J. Cristia, S. Cueto, O. Malamud, and Y. Cruz-Aguayo. 2015. One laptop per child at home: Short-term impacts from a randomized experiment in Peru. *American Economic Journal: Applied Economics*, 7(2), 53–80.

J. Cristia, P. Ibarraran, S. Cueto, A. Santiago, and E. Severin. 2012. Technology and child development: evidence from the one laptop per child program. SSRN Electronic Journal.

B. Piper, S. S. Zuilkowski, D. Kwayumba, and C. Strigel. 2016. Does technology improve reading outcomes? Comparing the effectiveness and cost-effectiveness of ICT interventions for early grade reading in Kenya. *International Journal of Educational Development*, 49, 204–214.

Janna Anderson, Lee Rainie and Alex Luchsinger. 2018. "Artificial Intelligence and the Future of Humans." Washington, DC: Pew Research Center.

R. R. Nadikattu. 2016. The emerging role of artificial intelligence in modern society. *International Journal of Creative Research Thoughts, 4(4), 906-911.*

Tomašev, N., Cornebise, J., Hutter, F. et al. AI for social good: unlocking the opportunity for positive impact. Nat Commun 11, 2468 (2020). https://doi.org/10.1038/s41467-020-15871-z

OECD. 2019. Artificial Intelligence in Society, OECD Publishing, Paris, France. Retrieved from https://doi.org/10.1787/eedfee77-en

Microsoft. 2018. The Future Computed: Artificial Intelligence and its role in society. Redmond, Washington: Microsoft Corporation.

R. Benjamin. 2019. Race After Technology: Shining Light on the New Jim Code. Cambridge, UK: Polity.

N. Byrnes. 2016. Why We Should Expect Algorithms to Be Biased. MIT Technology Review. Retrieved from https://www.technologyreview.com/s/601775/why-we-should-expect-algorithms-to-be-biased/

S. Noble. 2018. Algorithms of Oppression: How Search Engines Reinforce Racism. New York: NYU Press

National Science and Technology Council. 2016. The National Artificial Intelligence Research and Development Strategic Plan. Retrieved from https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf

M. Eslami, A. Rickman, K. Vaccaro, A. Aleyasen, A. Vuong, K. Karahalios, K. Hamilton, and C. Sandvig. 2015. I always assumed that I wasn't really that close to [her]: Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 153–162.

J. Okkonen and S. Kotilainen. 2019. Minors and Artificial Intelligence – implications to media literacy. Information Technology and Systems: Proceeding of ICITS 2019, 881-890. https://doi.org/10.1007/978-3-030-11890-7_82

J. Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society 3, 1, 2053951715622512.

Hitron, T., Orlev, Y., Wald, I., Shamir, A., Erel, H., & Zuckerman, O. (2019, May). Can children understand machine learning concepts? The effect of uncovering black boxes. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-11).

N. Diakopoulos. 2014. Algorithmic-Accountability: The investigation of Black Boxes. Tow Center for Digital Journalism, Columbia University. https://doi.org/10.7916/D8ZK5TW2

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, *38*(3), 50-57.

D. Long & B. Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376727

Druga, S., Williams, R., Park, H. W., & Breazeal, C. (2018, June). How smart are the smart toys? Children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 231-240).

G. Ptaszek. 2020. From Algorithmic Surveillance to Algorithmic Awareness: Media Education in the Context of New Media Economics and Invisible Technologies. In J. Ratajski, & M. Fedorowicz [Eds] *Media education as a challenge.* Warsaw: Academy of Fine Arts in Warsaw Publishing. Retrieved from https://www.academia.edu/43171356/From_Algorithmic_Surveillance_to_Algorithmic_Awareness_Media_Education_in_the_Context_of_New_Media_Economics_and_Invisible_Technologies_Democratisation_of_the_web_and_the_beginnings_of_data_driven_capitalism

AI4ALL. 2020. https://ai-4-all.org/

MIT Media Lab. 2020. Learning about artificial intelligence: A hub of MIT resources for K-12 students. Retrieved from https://www.media.mit.edu/articles/learning-about-artificial-intelligence-a-hub-of-mit-resources-for-k-12-students/

D. Touretzky, C. Gardner-McCune, F. Martin and D. Seehorn. 2019. Envisioning AI for K-12: What Should Every Child Know about AI? Proceedings from the Thirty-Third AAAI Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence.

G. Ptaszek. 2019. Media Education 3.0? How Big Data, Algorithms, and AI Redefine Media Education. In Divina Frau Meigs, Sirkku Kotilainen, Manisha Pathak-Shelat, Michael Hoechsmann, Stuart R. Poyntz [Eds] The Handbook on Media Education Research. New York: Wiley- Blackwell (International Academic Media Studies and Communication Global Handbooks in Media and Communication Research) [PDF] ttu.edu

Register, Y & A. Ko (2020). Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics. ACM International Computing Education Research Conference (ICER), 67–78. https://doi.org/10.1145/3372782.3406252

Freire, P. (1996). Pedagogy of the oppressed (revised). *New York: Continuum*.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33-35.

Mejias, S., Thompson, N., Sedas, R. M., Rosin, M., Soep, E., Peppler, K., ... & Bevan, B. (2021). The trouble with STEAM and why we use it anyway. *Science Education*, *105*(2), 209-231.

Maeda, J., & Burns, R. (2005). Creative code. *Education*, *7*, 177.

Peppler, K., & Kafai, Y. (2005). Creative coding: Programming for personal expression. *Retrieved August*, *30*(2008), 314.

Engel, D., & Phillips, J. (2019). Applying conservation ethics to the examination and treatment of software-and computer-based art. *Journal of the American Institute for Conservation*, *58*(3), 180-195.

Desai, D., & Darts, D. (2016). Interrupting everyday life: Public interventionist art as critical public pedagogy. *International Journal of Art & Design Education*, *35*(2), 183-195.

Habib, S. (2020). Art, social justice and critical pedagogy in educational research: The Portrait of an Artist as a Young Person. In *Handbook of Qualitative Research in Education*. Edward Elgar Publishing.

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.

Wright, D. E. (2020). Imagining a more just world: critical arts pedagogy and youth participatory action research. *International Journal of Qualitative Studies in Education*, 33(1), 32-49.

Wright, D. E. (2021). Countering Dominant Discourses about Youth: Critical Arts Approaches to Analyze Ideological and Institutional Oppressions. In *Engaging Youth in Critical Arts Pedagogies and Creative Research for Social Justice* (pp. 23-46). NY: Routledge.

Knochel, A. D., & Patton, R. M. (2015). If art education then critical digital making: Computational thinking and creative code. *Studies in Art Education*, *57*(1), 21-38.

C. Lee and A. Garcia. 2014. ""I Want Them to Feel the Fear…": Critical Computational Literacy as the New Multimodal Composition." In *Exploring multimodal composition and digital writing*, pp. 364-378. IGI Global.

E. Soep, C. Lee, S. Van Wart, T. Parikh. 2020. Code for What? In Jenkins, H., S. Shresthova, G. Peters-Lazaro (Eds.), *Popular Culture and the Civic Imagination: A casebook* (pp. 89-99). New York, NY: New York University Press. For more on the concept of Civic Imagination, please see https://www.civicimaginationproject.org/.

Gutiérrez, K. D., & Jurow, A. S. (2016). Social design experiments: Toward equity by design. *Journal of the Learning Sciences*, *25*(4), 565-598.

A. Strauss and J. Corbin. 1994 "Grounded theory methodology." Handbook of qualitative research 17, no. 1: 273-285.

E. Soep and V. Chávez. 2005. Youth Radio and the pedagogy of collegiality. *Harvard Educational Review*, 75(4), 409–434.

a.m. brown. 2017. Emergent Strategy: Shaping Change, Changing Worlds. Chico, CA. AK Press.

N. Frey and D. Fisher. 2013. Gradual Release of Responsibility Instructional. Framework. ASCD. Retrieved from https://pdo.ascd.org/lmscourses/pd13oc005/media/formativeassessmentandccswithelaliteracymod_3-reading3.pdf

M. K. Reha. 2019. How to Use "I Do – We Do – You Do" in Teaching. Grand Canyon University. Retrieved from https://www.gcu.edu/blog/engineering-technology/how-use-i-do-we-do-you-do-teaching

Gaslight. APA Dictionary of Psychology. 2020. Retrieved from the American Psychological Association's online dictionary.