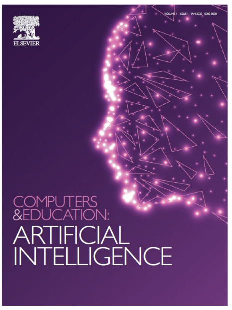


Co-designing AI with Youth Partners: Enabling Ideal Classroom Relationships through a Novel AI Relational Privacy Ethical Framework

Michael Alan Chang, Mike Tissenbaum, Thomas M. Philip, Sidney K. D'Mello

PII: S2666-920X(25)00004-9

DOI: https://doi.org/10.1016/j.caeai.2025.100364

Reference: CAEAI 100364

To appear in: *Computers and Education: Artificial Intelligence*

Received Date: 17 September 2024

Revised Date: 7 January 2025

Accepted Date: 7 January 2025

Please cite this article as: Chang M.A., Tissenbaum M., Philip T.M. & D'Mello S.K., Co-designing AI with Youth Partners: Enabling Ideal Classroom Relationships through a Novel AI Relational Privacy Ethical Framework *Computers and Education: Artificial Intelligence*, https://doi.org/10.1016/j.caeai.2025.100364.

Title: Co-designing AI with Youth Partners: Enabling Ideal Classroom Relationships through a Novel AI Relational Privacy Ethical Framework

Michael Alan Chang[a] (machang@bu.edu)
Mike Tissenbaum[b] (miketiss@illinois.edu)
Thomas M. Philip[d] (tmp@berkeley.edu)
Sidney K. D'Mello[c] (sidney.dmello@colorado.edu)

[a] Boston University
Two Silber Way, Boston, MA 02215
United States

[b] University of Illinois
1310 S. Sixth St. Champaign, IL 61820-6925
United States

[c] University of Colorado, Boulder
6661 CO-7, Boulder, CO 80303
United States

[d] University of California, Berkeley
2121 Berkeley Way, Berkeley, CA 94720
United States

Corresponding Author: Michael Alan Chang
Email: machang@bu.edu
Address: Boston University, Two Silber Way, Boston, MA 02215, United States

Declaration of Interest (for Title Page): This research was made possible by NSF grant DRL-2019806 and NSF grant #2021159.

Title: Co-designing AI with Youth Partners: Enabling Ideal Classroom Relationships through a Novel AI Relational Privacy Ethical Framework

Abstract: In recent years, the design of AI-based tools for educational spaces have been largely driven by researchers who impart their past expertises, experiences, and perspectives in the design process. While this typically leads to technically feasible designs and are often well-grounded in theories of learning, youth agency is typically limited in this process. In this paper, we argue that designers have a significant ethical responsibility to incorporate youth voices – in particular, their *dreams and concerns* – into the design of AI tools starting from conception. This need is particularly important as new applications for AI, such as AI-supported collaboration, introduce new surveillance vectors into classroom spaces. Drawing from recent scholarship which advances ethics and relationality in participatory co-design with youth, we introduce a co-design methodology in which youth are supported in imagining expansive technical possibilities for K-12 public schools, grounded within affordances, limitations, and tradeoffs of AI and machine learning techniques. This approach is demonstrated through our *Learning Futures Workshop*, which brought together 30 historically minoritized youth in conversation with experts in both education and technology. Through detailed case study on the enactment of the workshop, including a thematic analysis of the activities the youth engaged in and their outputs, we identified new, expansive relational possibilities for AI, ethical commitments to support the design, and finally, developed a novel AI Relational Privacy ethical framework that supports the design of new collaborative AI platforms. We conclude by connecting these findings and frameworks to the design of newly enacted AI-based applications and underlying data infrastructures.

Section 1: Introduction

While the use of educational technologies that can track student activity and products to help support student learning and orchestrate classroom activities has been around for decades (Andersen, Boyle & Reiser, 1985; Aslan et al., 2019), the rise of Artificial Intelligence (AI) and natural language processing (NLP) techniques has rapidly advanced the computational affordances of these systems. Tools that support collaboration is one area that has benefited from these developments, as they can now operate within authentic classroom contexts, record written, verbal, and non-verbal collaborations, and based on copious amounts of previously recorded data, automatically infer conversational semantics and dynamics (Adams et al., 2023; Stewart et al., 2023; Breideband et al., 2024). These tools can now be woven into the day-to-day minutiae of classroom activity, opening up a variety of exciting new learning interventions. At the same time, students in particular face a new and invasive surveillance vector; any utterance made during a collaborative learning activity—an inside joke whispered between friends, a private ask for help from a student to teacher—could all potentially be fodder for analysis and intervention by an AI tool. It can also be stored and shared beyond the small group to the teacher or researchers.

Due to the novelty and potential intrusiveness of these new collaborative AI interventions, designers of these tools have an obligation to anticipate the emergent ethical quandaries, particularly about students' educational data collection and how (if at all) that educational data is used to support innovative collaborative learning interventions. This issue of data governance has been central in the space of AI/Ed ethics, but the majority of those ethical discussions were situated in contexts where learning was envisioned to happen in more controlled, individualistic contexts: intelligent tutoring systems, automated assessment and feedback, etc. (Chiu et al., 2023, Heeg & Avraamidou, 2023). The stakes are particularly high when it comes to students of color; ethicists from sociocritical perspectives provide substantial evidence that data driven learning tools can, and have been, readily appropriated into logics of policing commonly found in many public educational contexts (Slade & Prinsloo, 2013; Tanksley, 2024). Yet, as Tanksley and others remind us, critical perspectives can inform design, in ways that serve potentially expansive ends while protecting non-dominant stakeholders from harm. Ultimately, a key component of ethical AI design is weighing possibilities and risks, and then continuously adapting design to the peculiarities of a context.

To anticipate this tension, we must meaningfully engage educational stakeholders – particularly students -- who will ultimately be left wrestling with these ethical questions and consequences of these decisions in actual pedagogical environments. Otherwise, designers of these tools risk building learning tools that are untrustworthy, inequitable, disposable, and ineffective (Lawrence et al., 2023; Ahn, et al., 2021; Ozmen Garibay et al., 2023; Rheu et al., 2021). Complicating matters is the general lack of approaches for effectively bringing in students as part of the co-design process of AI systems (Xu et al., 2023). Recent work has aimed to address this by including teachers in the co-design process (Lawrence et al., 2022); however the work with students as co-designers has been relatively limited (Van Brummelen et al., 2023), despite its value across many other areas of computer-supported learning (Druin, 2002; Ahn, 2021).

Here we aim to make two contributions: an approach to participatory design of collaborative AI with youth and a relational privacy ethical framework that emerged from our empirical results. To briefly elaborate, this paper aims to outline an approach to participatory design that centers historically minoritized students in the design process, with a focus on ethical co-design approaches that support the implementation of AI-supported collaboration in K-12 contexts. In the sections below, we outline a participatory design approach through which we sought to surface students' hopes for expansive collaborative possibilities with AI, which we called the Learning Futures Workshops. Next, we hone in on a key tension between possibilities and ethical issues surrounding data collected during collaborative learning. We then analyze workshop findings through the lens of contextual integrity (Nissenbaum, 2004) and based on those findings, identify three key ethical commitments to support the equitable design of AI-supported collaboration tools. To support designers in operationalizing those ethical commitments, we derive a novel ethical design framework called the AI Relational Privacy framework that can be more systematically used to guide design processes. Finally, to illustrate how to apply this ethical framework, we present an exemplar tool, the Community Builder.

Section 2: Background

Here, we introduce past literature that provides a basis for how to partner with youth to surface their hopes and concerns around the use of immersive, AI-supported collaboration tools. We begin by providing an overview of agentic, ethical, and relational approaches to co-design from human-centered design (HCD) and the learning sciences that have been employed to support youth in co-designing educational technologies. This body of work, covering the design of computing systems from conception to repair, tends not to explore youths' perspectives on AI ethics that come up in the design of these ed-tech systems. Thus our second background subsection provides an overview of AI ethics in education with a focus on data governance, and we explain why this body of work is insufficient to support our explorations of ethical issues that emerge from AI-supported classroom collaboration. To address this gap, our third and final subsection will introduce Nissenbaum's framework of contextual integrity (2004) and situate it in the context of AI-supported classroom collaboration.

Section 2.1: Relational Approaches to Co-design of Ed-Tech with Youth

Engaging youth in participatory design requires different approaches than designing with adults (Druin, 2002). Traditional power dynamics, notions of authority, and adults "knowing more" can cause friction in the design process if not properly attended to (Guha et al., 2013; Kumar et al., 2018). Without addressing this dynamic, technologies are prone to being designed as if children were "little adults", rather than existing with their own norms, needs, and concerns (Druin, 2002). At the same time, participatory design with youth can be particularly powerful as children are not as concerned about how things "should work" (Druin, 1999). However, many of these approaches often foreground the design process of the final product with less attention paid to youths' values, experiences, perspectives, and beliefs (Iriarte et al., 2023). For instance, product-oriented participatory design focuses on the design of an end product, which can limit the range of ideas and input that youth can provide (DiSalvo et al., 2017; Iriarte et al., 2023).

Specifically in the context of building educational technologies, participatory design has been uplifted as an approach to the design of educational technologies that recognizes that students are best served when they are given a high degree of agency in the tools that are developed for them (Druin, 1999). Placing students at the center of the design process mirrors the principles of learner-centered design that sits at the heart of much of the field of computer-supported collaborative learning (Bonsignore, 2013).

More recently, there has been a growing recognition that in order to effectively engage students in participatory design approaches, designers need to not just understand how students feel about the tool itself, but also apply human-centered design (HCD) approaches to better understand their personal values, dreams, concerns, and socio-emotional needs (Iriarte et al., 2023; Shehab et al., 2021). As educational technologies become increasingly integrated into all facets of students' learning experience, there are increased issues around students' perceptions of privacy, autonomy, and sense of security (Hourcade et al., 2016; Kumar et al., 2018). By applying an HCD lens to the participatory design process, designers can more effectively engage students in co-investigating the issues towards deeply understanding what matters to them, and design educational technologies that are more equitable and relational (Guha et al., 2013). Such an approach is particularly important when researchers and designers are working with students around technologies that they may only be tangentially familiar with, despite their potential impact in their lives and classrooms (Kumar et al., 2018).

One such “less familiar” technology is AI, which has experienced a meteoric rise in interest in recent years, to support classroom learning and collaboration (Touretzky et al., 2019). While there has been some work engaging teachers in the design of these systems and their related curricula (Chiu, 2021; Hrastinski et al., 2019; Lin & Van Brummelen, 2021; van Leeuwen et al., 2018), there has been relatively little work that engages students as partners in the design process (Holstein et al., 2019). This broad lack of engagement with students as participatory designers runs the risk of not attending to their many valid ethical and privacy related concerns about the integration of AI into their classrooms (Akgun & Greenhow, 2022; Slade et al., 2019). A failure to attend to the concerns of students around issues such as privacy in the design of these tools, has resulted in students rejecting their inclusion in their classrooms and the tools needing to be pulled (Ahn et al., 2021).

Here, we take the relatively unique step of bringing these perspectives on human-centered design of educational technology tools into conversation with ethical and relational approaches to co-design that emerged from the learning sciences.

Much like HCD approaches, youth are seen as occupying important positions in those contexts and the focus similarly is on shifting, or re-mediating (Gutierrez, 2009) youth positionality from a “delegitimized stakeholder” to a meaningful partner in design. Unlike HCD approaches which focus on re-mediating relationships primarily in the co-design context itself, the learning sciences approaches also centers the re-mediation of relationships in the context being designed for (Gutierrez & Jurow, 2016; Bang & Vossoughi, 2016). In applying this framework, we see that it is insufficient to just elevate the voices of youth in design; the co-design space must also support the youth in re-imagining relationships that happen within the institutional context being co-designed for (Philip et al., 2023; Chang et al., 2024). In this case, our focal context is schools. Concretely, the space of exploration shifts from: “how can I as a researcher support youth in feeling authoritative in designing AI” to “how can I as a researcher support ideal relationships in classrooms, and what possibilities for AI does that open up?” In our approach, we bring participatory design research into conversation with aforementioned literature on co-designing AI with youth; we ask what happens when designers simultaneously support youth in imagining possibilities spanning both schools and artificial intelligence?

Moreover, what ethical considerations emerge as explore those expansive possibilities? In the next subsection, we situate major ethical considerations for the design of AI tools for collaboration in the broader context of inequitable classroom relationships.

Section 2.2: Ethical Considerations for the design of AI Tools for Collaboration

AI applied to educational contexts has raised many ethical questions at various stages of design, from conception to implementation to maintenance. Researchers in recent years have taken steps to study the ethics of AI and the integration of AI into classroom practices (Holmes et al., 2021; Nguyen et al., 2022). This analysis has raised a bevy of important questions around the nuances of inclusiveness, fairness, accountability, and transparency (e.g., black-box predictions made by AI models that disproportionately adversely affect people of color), and their integration with classroom practices such as pedagogies and teachers' decision making (Holmes et al., 2021; Office of Educational Technology, 2023). Overall, researchers in recent years have made a compelling case that the ethics of AI-Ed cannot be considered independently of an analysis of the social-relational context that those AI tools are deployed within (Ifenthaler & Schumacher, 2016; Kitto & Knight, 2019).

Perhaps the most prominently discussed issue is the ethics of collecting and handling educational data, a fundamental requirement to training and leveraging any machine learning model. Learning analytics researchers, despite not necessarily using AI in their approaches, have made fundamental contributions here, as they explore the nuances of ethically attending to data interpretation, informed consent and privacy, and the politics of data storage (Pardo & Siemens, 2014; Prinsloo & Slade, 2015; Draschler & Geller, 2016; Pargman & McGrath, 2021). Across these framings, as illustrated by Nguyen et al. (2023), student consent over collection of their individual learning data is a central theme although the object of that consent can vary substantially. For instance, Corrin et al. (2019) suggests that even if consent is given by students, designers are ethically obligated to consider students' intentions for how long to store the data and who can access it. Others, such as Li (2007), urge designers to consider the unexpected, possibly invasive predictions that students' personal data can be used much later down the line.

These issues of data privacy are intertwined with other ethical issues in schooling contexts in general. For instance, Slade & Prinsloo (2013) sensitize designers to existing hierarchical relationships in schools that are used to justify the surveillance of students. In creating more ethical systems, they caution designers against making sweeping generalizations; instead, designers should attune their processes to the context and dynamics that shape student identity, preferences, and values.

In a similar vein, others have argued that ethical questions about AI and data governance are intimately connected with racial inequity. Researchers have additionally shown how AI tools can mediate racism and anti-blackness in schools reproducing broader racial logics within the context of K-12 schools (Tanksley, 2024). For black and brown youth, who have historically been under more surveillance scrutiny, AI tools that collect data about classroom dynamics can raise additional complexities in how they would welcome those tools into the classroom. Yet these researchers encourage designers to not only consider how to address the racism embedded in *how* an AI tool is developed, but also to encourage designers to "rebuild and reimagine" new technologies towards ideal, liberatory futures. In this work and others outside of AI-specific contexts, we are reminded that ethical and responsible design is not only about anticipating concerns and issues, but also creating the opportunities for imagining new possibilities (Stilgoe et al., 2013) – ones that move the needle to educational (and beyond) futures that transform the inequitable status quo.

Overall, past work on AI/Ed ethics highlights the intricate dance designers play as they move between possibilities, design approaches, and integration with classroom practices, as they build towards ethical and equitable futures. Building on this work, our specific context of AI-supported collaboration raises several unique considerations. Firstly, as described in earlier sections, little is known about students' hopes for an AI agent's ideal role in collaboration. Secondly, the ethical questions surrounding data privacy and the integration of AI into classroom practices are particularly urgent. In contrast to the majority of AIEd research which focus on individualized learning systems operating within relatively simple relational context (e.g., a student working individually with an intelligent tutoring system that shares data with a teacher), our focus on a collaborative AI agent requires us to grapple with how the tool intervenes within multiple levels of activity and discourse operating within the ecosystem of a classroom. For instance, in a typical small-group collaboration, there may be whole-group level discussions with a teacher, a whole-group discussion without a teacher, two friends making a private inside joke, a teacher whispering in one student's ear, just to name a few. Should an AI agent be privy to all these discussions? The specific context of AI-supported collaboration raises a set of ethical issues that have not previously been explored in AI-Ed ethics settings. To help us answer these questions, we look towards relational approaches to data privacy that have not been widely adopted in educational research.

Section 2.3: Relational Approaches to Data Privacy

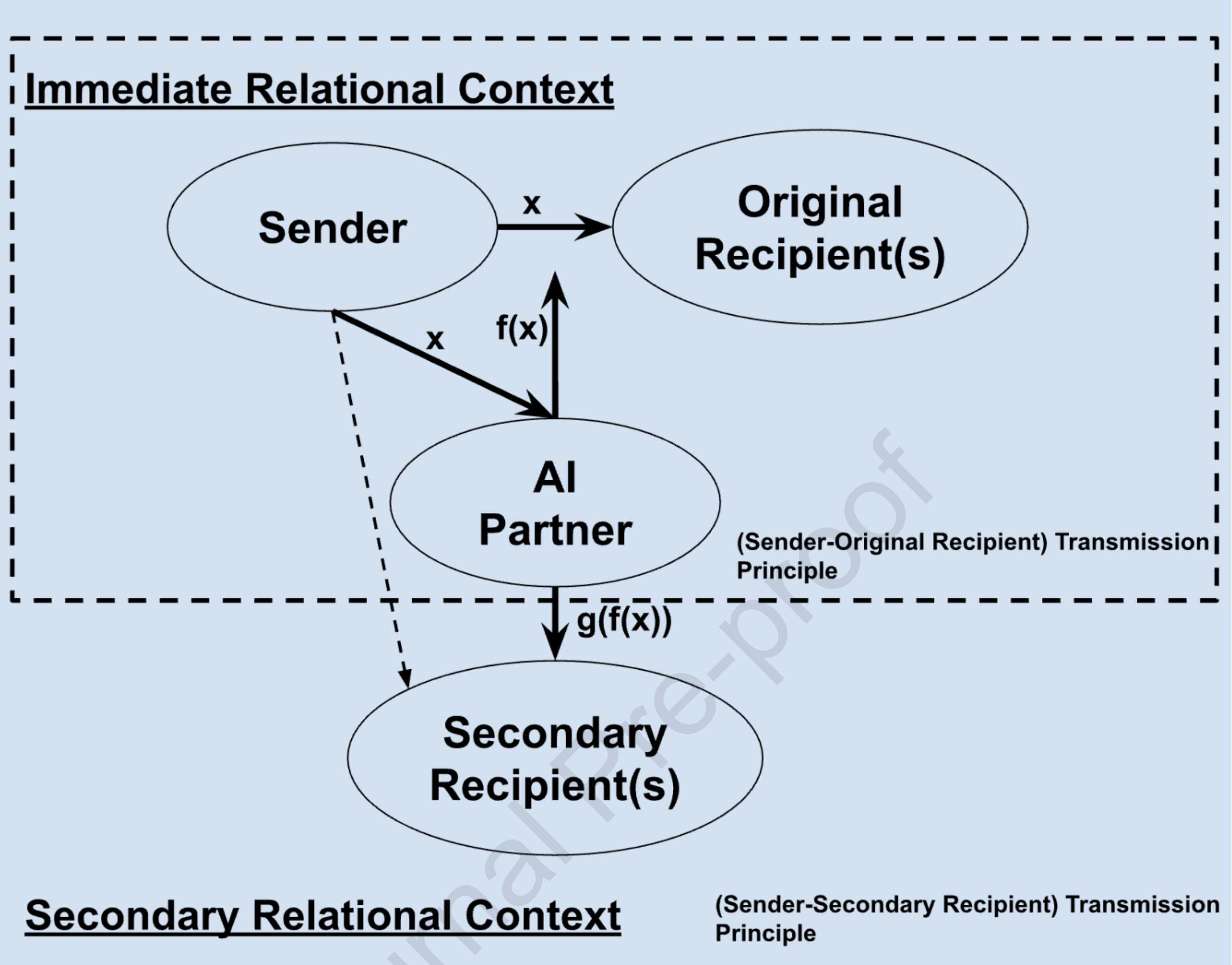
To help us develop an ethical framework for data privacy in AI-supported collaboration, we look to Nissenbaum's theory of contextual integrity (2004). Critical to Nissenbaum's framework is the notion that all information flow occurs within some social context (e.g., education, healthcare, religion, etc.), marked by specific practices, goals, and ends; participants within these contexts are aware they are participating in this practice and can explicate the norms of this context. For the purpose of modeling information transmission within these contexts, the context is parameterized by four categories: sender, recipient, information "types," and transmission principles. Every agent operating inside a context has a role (e.g., student, teacher, administrator). Transmission principles represent terms and conditions that dictate how information is shared between participants (consent, coerced, in confidence, via a warrant, etc.). Contextual integrity is met when 1.) information types are appropriate for a given context; and

2.) the flow of information follows the transmission principles. If contextual integrity is not met, user's expectations of privacy are violated, provoking public outcry, leading to lack of trust and adoption towards the intervention.

Contextual integrity has primarily been used to support designers in ethical reasoning at the organizational level rather than at the activity level where these information flows are actually playing out. Nevertheless, contextual integrity offers a designer-friendly formalization of our privacy within complex relational spaces, while at the same time making salient the relational possibilities and concerns, which featured so prominently in our findings. While the static notion of "roles" can be a stable and useful construct, the educational classroom context in particular demands a more flexible understanding of these norms, where personal and academic boundaries are constantly being breached and re-assessed. As Criado and Such (2015) argue, context, roles and associated informational norms are "implicit, ever changing and not a-priori known."

Figure 1

Transfer of data recorded by an AI partner in a classroom, through the lens of contextual integrity


Immediate Relational Context
Sender
x
Original Recipient(s)
x
f(x)
AI Partner
(Sender-Original Recipient) Transmission Principle
g(f(x))
Secondary Recipient(s)
Secondary Relational Context
(Sender-Secondary Recipient) Transmission Principle

Bringing this discussion specifically into our goals of supporting classroom collaboration, Figure 1 shows how contextual integrity frames one model of AI-supported collaboration. At the most immediate level, a sender is conveying some information x to the original recipient(s), supported by an AI partner recording the conversation. The shared information is subject to particular expectations of privacy, based on the following contextual integrity tuple: (sender, original receiver(s), the information type of x, the (sender-original recipient) transmission principle). Additionally, the AI collaboration partner also collects x and renders some computation over it, the result of which we refer to as f(x). For instance, x may be the recorded utterances of collaborators, while f(x) may be the output of a supervised classifier that determines whether an utterance is an effective collaborative move.

The AI partner may then do two things with f(x). First, f(x) may be shared directly by the AI partner back to the sender and original recipient, e.g., an AI partner providing feedback to collaborators about their collaborative efforts. Second, f(x) may also be shared with a secondary recipient (e.g., a teacher who was not part of the original discussion), and may be subject to additional computational transformations which we refer to as g(f(x)). For instance, g(f(x)) may strip personally identifiable information from f(x) so that collaborator privacy is preserved. The expectations for privacy on this second exchange depend on the following tuple: (sender, secondary receiver(s), the information type of x, the (sender-secondary recipient) transmission principle). At the highest level, the central tension that we identified are the conditions for passing of information x between the immediate relational context and the secondary relational context.

This model provides a means for designers to reason about the nuances of consent that emerge when multiple learning contexts overlap, as has been extensively studied by those who bring a sociocultural perspective to collaborative learning. Gutierrez et al., (1995) view learning happening at three relational context levels: a first space (traditionally formal, “on-task/on-topic” academic talk and activities), a second space (the classroom “underlife” that the teacher is not supposed to be aware of), and a third space (a space between the first and third space that a politically-aware teacher skillfully navigates between). Under Gutierrez’s definition, classrooms always have a first space and a second space, each with presumably different expectations of privacy. The third space, on the other hand, is under collective negotiation between classroom actors. For instance, we might envision two students privately talking with each other in the second space; in our Figure 1 framework, this would be information exchanged in an immediate relational context. However, if an AI collaborative agent were to share that discussion with a teacher, that would violate the students’ expectations of privacy for a second space conversation. Ethical questions emerge, then, about under what conditions it would be appropriate to share such a discussion with a teacher without violating the trust of the students. If the conditions of a classroom are supportive of third spaces, would it then be appropriate to share such a discussion with a teacher?

We engage more deeply with this framework of conceptual integrity in the following sections, and explore how youth navigate the grey areas introduced by the insertion of a collaborative AI tool. In the sections below we show how this model can be applied to understand how students, engaged in a codesign workshop, engage and consider the use of collaborative AI agents in their classrooms.

Section 3: Material and Methods

We start by describing the author’s institutional context to explain some of the constraints that shaped the design of our workshop. Next, we describe the approach of the Learning Futures Workshop and provide concrete details about the workshop implementation and participants. Finally, we describe our data and analysis processes.

Section 3.1: Author’s Institutional Context

The authors of this manuscript are members of an interdisciplinary AI Institute called the Institute for Student-AI Teaming (iSAT). iSAT was founded to develop AI-based tools that support small-group collaborations in United States K-12 classrooms (D’Mello et al., 2024). iSAT researchers share a commitment to Responsible Innovation (Stilgoe et al., 2013). Responsible Innovation orients us to *inclusively anticipate* concerns and dreams held by impacted actors, and also to be *responsive* to those dreams and concerns in our design efforts. Early on, iSAT technical and educational experts stated a commitment to being responsive to the findings in our Learning Futures Workshop; thus as designers of the workshop, we sought to walk a line between (a) youth's expansive hopes and dreams for schools which may extend beyond iSAT's implementation capabilities and (b) the existing technical and educational expertises that were featured in iSAT. Throughout our workshop, we explicitly communicated this tension to youth and made the Institutional context transparent.

Section 3.2: Workshop Approach

Our Learning Futures Workshop sought to surface youth's hopes for expansive collaborative possibilities inside schools by re-mediating (Gutierrez et al., 2009) key relationships across *both* schools and AI. This focus on re-mediation (instead of remediation) supports us in moving past deficit-oriented views of youth towards understanding and re-imagining the complex ecologies where youth occupy. Given the constraints described in Section 3.1, we surfaced and *re-mediated* the following key relationships:

Our first focus was the relationship between *technology designers* and *youth,* where youth expertise is commonly dismissed. We rectify this by positioning youth in our workshop space as *experts* in AI and schools, who have opportunities to freely imagine, propose, and create expansive AI possibilities. Facilitators specifically took a step back, primarily serving to facilitate discussion, and occasionally contributing technical expertise when appropriate to create a greater sense of feasibility around a technology-based proposal

Second, we re-mediated the relationship between *developers of Artificial Intelligence* (AI) and *users of the tools.* At the time this workshop was held, iSAT primarily was interested in investigating empirical-based AI approaches, which relied on collecting large volumes of user data, and making inferences over that data using black box statistical approaches. Extensive data collection is essential to the effectiveness of these approaches, and has been shown to dovetail with broader systems of authority and surveillance, particularly around oppressed populations (Benjamin, 2019). Moreover, these largely black box models have been shown to take on harmful biases against minoritized groups. Thus, when positioning youth as developers of these technologies, we emphasized that youth had the ultimate say about (a) what data should and should *not* be collected about them and (b) what should and should *not* be inferred from that data.

Thirdly, we re-mediated the relationship between students and other actors (e.g., other students, teachers, and community) in American K-12 public school collaborative contexts, a context alike our iSAT's existing district partnerships. Under the grammar of schooling (Tyack & Tobin, 1994) that has defined the American social imagination of public school education, collaboration is often oriented towards narratives of efficiency, driven and monitored by an authoritative teacher or high-status classroom actors. Compounding this challenge: even outside the explicit space of collaboration, school structures often delegitimize youth as stakeholders in public schools broadly as changemakers (Brion-Meisels & Alter, 2018). Thus when positioning youth as agents of relational change in classrooms, we made clear that youth's classroom experiences were of foremost importance in the dreaming of new worlds, and created opportunities to dream about their ideal classrooms alongside technological possibilities.

Section 2.3: The Learning Futures Workshop Implementation

Our Learning Futures Workshop occurred remotely over five days during the summer of 2021, for three hours each day. To prepare for the dreaming phases of our workshop, the first two days of our workshop sought to highlight the key *affordances* and *limitations* of supervised, empirical machine learning techniques, particularly applied towards natural language processing tasks. Here, we describe the early workshop activities but, for brevity, do not go into great detail about youth's engagement. For more details about that process, see our previous writings about the workshop empirical findings (Chang et al., 2022).

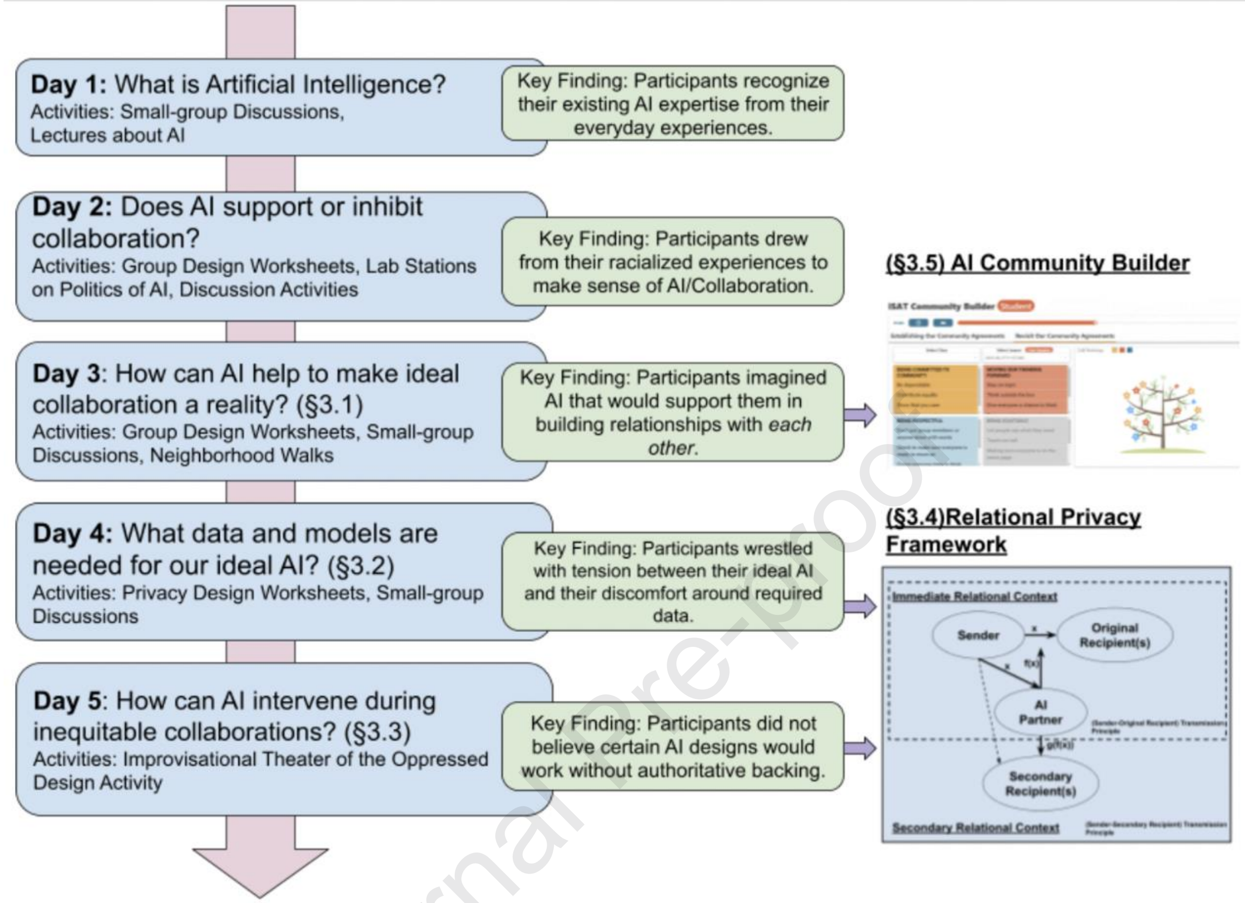
In the opening activities of our workshop, we asked youth: "where do you experience AI in your everyday life?" Many surfaced their experiences using recommendation algorithms in social media sites (e.g., Instagram, TikTok). Building on youth's personal experiences, we provided a brief lecture explaining the high-level mechanisms of supervised, empirical machine learning techniques. In particular, we broke down those techniques into *training* and *inference*. Training identifies patterns over (typically) large data sets, while inference uses previously trained models to categorize previously unseen data. We emphasized that AI models could *infer* unexpected patterns from the data, e.g., a retail store identifying a customer as pregnant from her purchasing activity.

After the lecture, facilitators led youth in an activity where they could design an AI algorithm that could support collaboration in social media sites. This activity offered us a chance to start making connections to small-group collaboration, our workshop's focal classroom context. To scaffold this process, facilitators started by asking youth what AI actions and predictions might support them in collaborating. Then, facilitators asked youth to consider how raw data collected by social media sites today (e.g., likes, clicks, etc.) might inform predictive outcomes. In the final activity on day 2, facilitators introduced a sociopolitical perspective to AI and argued that technology takes on broader biases that exist in society. Youth then took part in a series of small group discussions, where they applied this newly learned lens towards a number of AI-driven applications: automated game moderation, a legal case involving AI and protected classes, Microsoft's Twitterbot Tay, and the short-lived deployment of police robot "digidogs" in New York City.

This brief introduction to AI paved the way for the last three days of the workshop, which were spaces of dreaming, designing, and enacting. Details of these activities are sequentially embedded in the following Results section.

Figure 2

Workshop Throughline



Day 1: What is Artificial Intelligence?
Activities: Small-group Discussions, Lectures about AI

Key Finding: Participants recognize their existing AI expertise from their everyday experiences.

Day 2: Does AI support or inhibit collaboration?
Activities: Group Design Worksheets, Lab Stations on Politics of AI, Discussion Activities

Key Finding: Participants drew from their racialized experiences to make sense of AI/Collaboration.

Day 3: How can AI help to make ideal collaboration a reality? (§3.1)
Activities: Group Design Worksheets, Small-group Discussions, Neighborhood Walks

Key Finding: Participants imagined AI that would support them in building relationships with *each other*.

Day 4: What data and models are needed for our ideal AI? (§3.2)
Activities: Privacy Design Worksheets, Small-group Discussions

Key Finding: Participants wrestled with tension between their ideal AI and their discomfort around required data.

Day 5: How can AI intervene during inequitable collaborations? (§3.3)
Activities: Improvisational Theater of the Oppressed Design Activity

Key Finding: Participants did not believe certain AI designs would work without authoritative backing.

(§3.5) AI Community Builder

(§3.4)Relational Privacy Framework

ISAT Community Builder
Student
Immediate Relational Context
Sender
x
Original Recipient(s)
f(x)
AI Partner
(Sender-Original Recipient) Transmission Principle
g(f(x))
Secondary Recipient(s)
Secondary Relational Context
(Sender-Secondary Recipient) Transmission Principle

Note. We summarize the key activities and findings of our workshop and show how it informs the design of the tools and frameworks described in later sections.

Section 3.4: Participants and Data Collection

Our Learning Futures Workshop worked with thirty youth from California, Colorado, and Oklahoma. All students were high-school aged, ranging from grades 9 to 12. 14 youth identified as male, 14 identified as female, and 2 identified as non-binary. Twelve self-identified as Asian-American or Pacific Islander, seven self-identified as Latinx, six self-identified as African American, two self-identified as Native American, and two self-identified as white. Zoom video recordings were saved for large-group and small-group discussions. Additionally, artifacts (worksheets and drawings completed using Jamboard) created during the workshop were saved.

Section 3.5: Analysis

Between each day of the workshop, workshop facilitators met and recorded shared themes and wonderings. Following the workshop, four members of the research team individually and collectively watched video recordings together; a key question emerging from these discussions were the peculiarities of using AI specifically within the interpersonal, collaborative context — especially around issues of data privacy. Based on this discussion, the first author conducted an inductive analysis, specifically observing what types of classroom relationships youth attempted to re-imagine. The first author initially transcribed the video recordings line-by-line, and then coded each turn where youth either expressed a hope that they hoped that AI could help with, or a concrete proposal involving AI in classrooms. Over initial dreaming activities in the workshop, we identified four relational contexts that youth participants sought to re-imagine: student-student, student-teacher, student-administrator, and student-community.

During this coding process, the first author wrote an analytic memo detailing how youth participants often blended these different collaborative contexts, particularly as the workshop advanced past conceptualization of new AI-collaboration metaphors into the implementation and evaluation of their proposed ideas. For instance, youth made an initial proposal where the AI helps to connect groups based on topical interest, a *student-student* relationship. However, when considering the data required by the AI to make it work, youth made salient *student-community* connections since their topical interests were influenced by community events. Additionally, the first and third author realized that these connections created tension because of different expectations of privacy over the salient contexts; the first and third author then deductively created codes based on the framework of contextual integrity (Nissenbaum, 2004) based on the the types of data being shared and the expectations of privacy (i.e., transmission principles) youth described for that data.

At the next level of coding, the first author coded each of the previously identified relational contexts with their contextual integrity parameters. Finally, as a research team, we discussed when youth participants drew connections between different relational contexts and identified contextual integrity violations that occurred when those contexts were brought together. Building on our research questions and our coding process, we present our thematic analysis as a detailed case study. Because of this shift mid-workshop from conceptualization to implementation, we present each day of the workshop sequentially and highlight identified themes along the way.

Section 4: Results

Prior to the activities described here in this Results, youth had learned about machine learning and a sociopolitical perspective towards technology. For each major activity in days 3-5 of the Learning Futures Workshop, we describe the design activity before diving into one representative vignette. We then show how these Results lead to the development of a novel Relational Privacy framework, and finish the section by showing how the Relational Privacy framework was operationalized in a novel AI tool built by our research institute iSAT.

Section 4.1: Ideal Schools and Imagining Possibilities (Workshop Day 3)

In small breakout groups, youth were given the following prompt: "what does ideal collaboration look like, and how might AI help to make that a reality in your schools today?" As the first dreaming activity, the focus was not on technical feasibility, merely the *ideation* of exciting proposals. In this vein, we did not view the AI proposals made in this activity as final or above criticism. Instead, in our analysis, we found it meaningful to look beyond the immediate proposals and understand where youth were coming from as they constructed these tentative proposals. In this vignette, we illustrate one breakout group's discussion around several AI possibilities. Participant quotes in this section are slightly modified to remove filler words.

In this small group, one participant Eric provided some framing to begin the discussion: "Okay, so my thought about this is not really technology, it's theory. Collaboration should be about equity not equality. In collaboration everyone should aim to give as much as they can and get as much as they can." Eric made central equity as a key goal of AI-supported collaboration, and more specifically, and emphasized that each student, depending on their circumstances, could "give" and "get" from each other in different amounts. This framing immediately implicates AI's role in mediating relationships *between students*, rather than just between the AI and an individual student.

Following Eric's remark, Ricky declared that teachers "can only help so many students…if you had AI, it could tailor itself to every student's needs." Ricky did not counter Eric's equity framing, but offered a different imagination about how AI supports equitable goals: the AI acts as a proxy for a teacher and individually supports students. While Ricky advocated for a student-AI tool, his framing of the tool ("teachers can only help so many students") suggests that the proposal addressed a key frustration, namely that teachers *could* support the youth at a personal level, but they lack the resources to give sufficient attention to students. We might begin to assume the relational context that Ricky describes, one where students work with AI, but with the expectation that teachers will have access to those conversations; therefore, an AI that were to share this information with a teacher would not violate contextual integrity.

Another student, Akil invoked a number of different possibilities and built on Eric's ideas. Akil first considered how AI tools could keep student peers from getting in "trouble for talking to your friend or asking a question." We might infer that Akil is speaking about a specific situation where a teacher is perhaps giving a lecture and silence is expected from students. While Akil's follow-up proposal operates in the same relational context as Eric's earlier proposal (a student-AI relationship), he specifically describes an expectation that the teacher should not be aware of the question he asked to the AI. Next, and in a similar vein, Akil discussed how "Google or Siri" could help answer a "question you weren't sure how to work or were embarrassed to ask." Finally, he imagined a proposal that brought the community into the classroom and capitalized on his community passions: "an AI that could connect you to different resources. Instead of googling jobs in my neighborhood, a career test! Where you excel, from there you do really well in these areas, we think these organizations would go well with you."

Esperanza later chimed in: "I feel like collaboration shouldn't be dreadful. In my classroom, it's dreadful many times." Later, she added: "I feel like people take [classes] to get credit, some people don't take it really seriously. I think it's a problem… I just feel like some people don't take it seriously in general. The whole class, the teacher, the topic!" After some additional prompting by the facilitator, Esperanza replied hesitantly with an AI that re-imagined student-student interactions: "Maybe the ones who are more experienced in it or more interested in it could somehow motivate the others to be interested in it as well."

Over the course of this discussion, youth surfaced a variety of frustrations and hopes for what collaboration in school could be like, spanning relationships with each other, with teachers, and with the broader community. Our analysis in this section reveals that ultimately, youth had deep, substantive hopes for re-mediating key relationships they experienced inside classrooms. Our co-design space created the opportunity for youth to imagine how they might tackle some of these frustrations and hopes using AI. While these proposals may raise many concerns about technical feasibility and learning theory, the discussions underlying the proposal makes clear that if we are to realize youth expansive hopes using artificial intelligence, we must better understand how AI can help support, accompany, and uplift *existing* classroom relationships.

While the focus of this activity was to imagine new possibilities for AI-supported collaboration, our lens of contextual integrity also begins to show how privacy expectations within each of these relational contexts are parametrized by more than just the type of data. In other words, as illustrated by the different motivations for Akil and Ricky's proposal, Ricky and Akil may ask the exact same question to an AI, but have entirely different expectations of privacy depending on the classroom context and the activity that they are asking those questions in.

In other groups, youth described wanting to feel cared for, respected, and heard in classrooms. Common AI proposals made in response to these experiences included a friendship finder ("Tinder but for friends") and a collaboration matcher based on shared interests. Another common frustration was youth's frustration with other youth for perceived disruptions to their collaborative efforts; they imagined an AI that would keep those disruptive youth accountable. These proposals formed a starting point for development and critique in the later days of the workshop.

Section 4.2: Designing AI, Considering Data (Workshop Day 4)

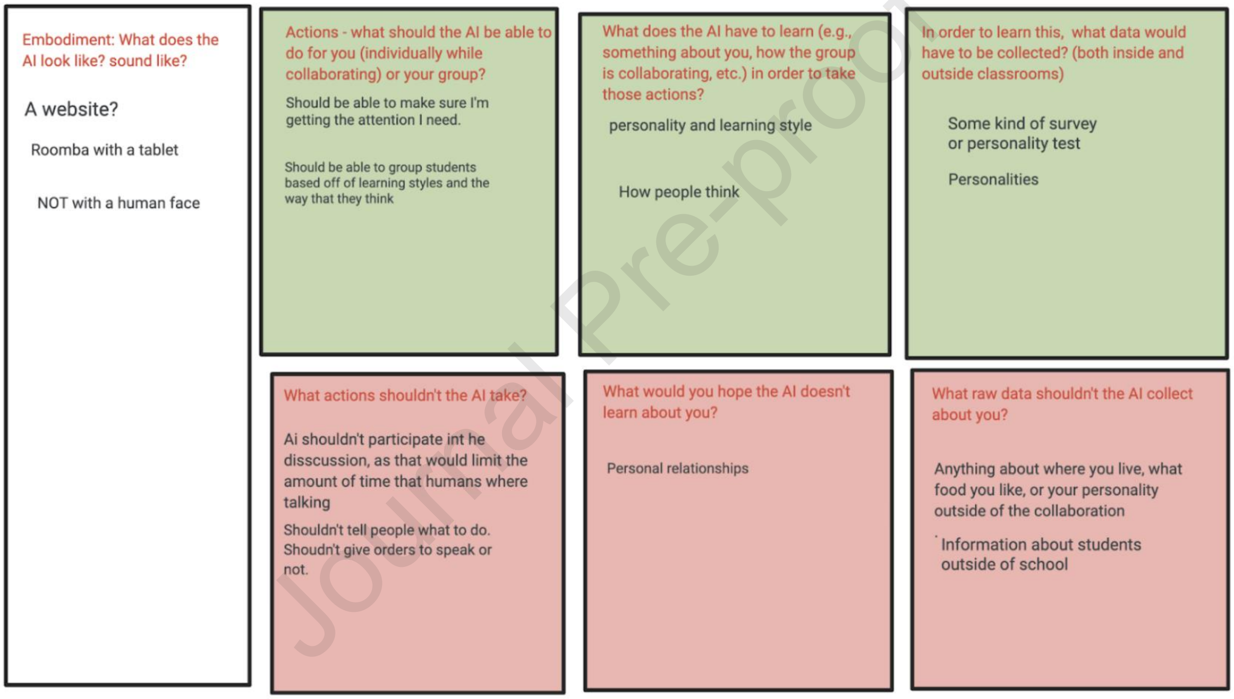
During day 4, we planned to take previously conceptualized AI proposals and better understand their ethical implications, particularly around an essential part of empirical AI systems: training models on large amounts of user data. The discussion was supported by a worksheet (Figure 3), which was custom created for the purpose of this exercise. The worksheet asked youth to collectively consider their proposed AI agent's embodiment, the actions taken by their imagined agent, the corresponding inferences the AI would need to make, as well as the raw data necessary to make those inferences. Additionally, for each of those dimensions, the worksheet explicitly asked youth to describe what they hoped the AI agent would *not* do.

In coding these excerpts, we observed that the worksheet raised new considerations for previously proposed AI features. These considerations were driven by youth's experiences around classroom relationships that *differed* from the specific relationship being re-imagined. We coded for these different relationships, and sought to understand how these newly salient relationships shaped how the young folks viewed their original relational possibilities for the AI.

In the following vignette, we share the discussion around one oft-proposed metaphor, which we have termed the Collaboration Matcher, an AI that supports youth in identifying collaborative partners. Initially, the Collaboration Matcher was designed to help mitigate the feeling of dread when it came to collaborating with *each other*: the focus was thus on the relationships between students and other students.

Figure 3

A completed data privacy worksheet used by youth to imagine the privacy tradeoffs involved in proposed AI tools



Embodiment: What does the AI look like? sound like?

A website?

Roomba with a tablet

NOT with a human face

Actions - what should the AI be able to do for you (individually while collaborating) or your group?

Should be able to make sure I'm getting the attention I need.

Should be able to group students based off of learning styles and the way that they think

What does the AI have to learn (e.g., something about you, how the group is collaborating, etc.) in order to take those actions?

personality and learning style

How people think

In order to learn this, what data would have to be collected? (both inside and outside classrooms)

Some kind of survey or personality test

Personalities

What actions shouldn't the AI take?

Ai shouldn't participate int he disscussion, as that would limit the amount of time that humans where talking

Shouldn't tell people what to do. Shoudn't give orders to speak or not.

What would you hope the AI doesn't learn about you?

Personal relationships

What raw data shouldn't the AI collect about you?

Anything about where you live, what food you like, or your personality outside of the collaboration

Information about students outside of school

In this breakout group, youth explored how an AI could identify how individuals learn by examining their individual attributes ("hav[ing] similar personalities") or learning styles ("some learn by reading, doing some activity"). These initial ideas, building on notions of "learning styles" that remain common in popular discourse even though they have been debunked in educational research (Newton, 2015), were later problematized by the workshop participants during this discussion; we omit those discussions for brevity.

Youth then collectively filled out the data privacy worksheet for the Collaboration Matcher. While filling out the worksheet, the participants stated explicitly that they did not feel comfortable having an educational AI agent collect personal data outside of the physical school boundaries. Picking up on this initial boundary, the facilitator seeded a discussion by asking, "How is the AI supposed to learn your learning style without that [out of school] information though?" Melinda quickly responded, "I don't think it needs to know our *personal* relationships with people.” Another participant, Sam, further refined the boundary:“I don't want them to know who I’m going out with.” From these excerpts, we don’t know the specific relational experience motivating Sam and Melinda’s discomfort with sharing personal relationships, but we infer that they are speaking to a relational dynamic where that data may be used to harm or embarrass them. Sam and Melinda began to define the transmission principles for information about personal relationships at the level of whole schools. The facilitator (a former teacher), pointed out that such a boundary around using personal data might constrain the capabilities of the Collaboration Matcher, arguing that teachers often observe romantic crushes between youth, and assign them into the same group in order to motivate them. Melinda affirmed the facilitator: “Teachers assign this because the boo know that they might be trying to impress them.” The Collaboration Matcher was initially designed to support student-student relationships; because we engaged the youth in considering the *data implications* of the AI tool, the relationship between student and teacher also became salient and led to certain relational possibilities for the Collaboration Matcher being closed off.

Sam then made an argument that the Collaboration Matcher could work even without personal information: “Compatibility has nothing to do with the level of friendship! I can never be in a group with [my friends]. I would never get anything done unless it’s an art project. You put me in a history group with my homies, F. Terrible! You put me with a random dude who I said wassup to a couple times in the corner of the room, we ace-ing that.” While the initial part of Sommer’s argument seemed to de-emphasize the importance of personal relationships, Sommer’s later argument in fact illustrated how personal relationships do indeed contribute to the effectiveness of the Collaboration Matcher. Faced with a relatively new set of considerations around a relationship-oriented AI metaphor, Sam and Melinda were actively making sense of tradeoffs and tensions.

This sense-making around boundaries continued. During a lull in the conversation, Sam wondered: “What if you are tardy [in class] for health reasons? Do you want the AI to know that? A lot of people would see that as an invasion of privacy. I would be okay with that. A recurring issue.” Sam acknowledged that others might not share his comfort, but for *this* specific application*,* he would give away private health data. Melinda, reflecting on Sam’s comment, later stated: “Schools have a lot of our information already. Would it just have all the information that the school already has? I have therapy, and the school already knows that. Still kind of sketch – but still!” Sam and Melinda continued to break away from the hard line between personal *and* academic data for AI use in classrooms. They both acknowledged the sensitivity of health information, but nevertheless deemed the tradeoff to be worthwhile. Now considering the student-administrator relationship, the lack of perceived empathy towards youth’s health conditions was sufficient to outweigh the potential harm from sharing that data. Sam and Melinda further nuanced the seemingly rigid transmission principles around personal data and health data that they drew at the outset.

Similar discussions occurred in other Day 4 breakouts in the workshop, where youth grappled with the tradeoffs of designing an AI that supports them in building relationships with *each other*. In other groups, Nick and Eno discussed sharing their individual "weaknesses," Juan and Kory contemplated biological signals like their "brain waves," and Larry wondered about out-of-school interests ("relate certain subjects to things out of schools"). Across these discussions, several findings become clear: designing relationships with AI lead to an unfamiliar and slightly uneasy set of discussions around *what data is collected* and *what could be inferred from that data, and who has access to any of these values.*

Taken together, youth's relational hopes for an AI agent (e.g., Collaboration Matcher) may not cleanly align with the relational data necessary to make those AI agents a reality. Relationships outside of the relationship under design become salient and close off or open up possibilities. Moreover, these expectations are constantly under negotiation; even over the course of our 45 minute discussion, youth increasingly nuanced an initially hard boundary between personal data and "academic" data.

Section 4.3: Theatro Activity (Workshop Day 5)

Our last day of the workshop involved a modified Theatro Activity (Boal & McBride, 2013). In Theatro, actors first perform a scenario where a perceived injustice occurs. After an initial performance, the actors repeat the scene but give way to audience members who "tap into" the scene and experiment with different interventions. To better understand the role of AI in collaboration, we designed the activity such that scenes must feature a virtual AI assistant (e.g., Amazon Alexa) and audience members could *only* step up to replace this virtual assistant. The Theatro script was co-created with workshop participants, who incorporated immediate needs they experienced while collaborating in schools. Through this activity, youth were able to test the boundaries of their comfort with an AI agent's capabilities (Holstein et al., 2017) and determine how an AI feature could effectively carry out its mission.

We highlight one small-group discussion, where youth experimented with an AI that kept "disruptive" youth from interfering with classroom collaborative efforts. This AI proposal was controversial, as one of the most commonly proposed AI tools that was also frequently complicated by youth. In the breakout, youth iterated through the script four times, trying out increasingly punitive measures to augment the AI's authority. These proposals ranged from threatening "disruptive" students by deducting grades, summoning the teacher, and finally, in the most extreme case, summoning parents. The proposal – which started off as a student-student relational tool – evolved to bring in teachers and parents. After this activity, participants debriefed their experience.

Eric initiated the conversation: "For it to be more effective, it would need more power, the ability to mute someone. I don't know what other power it would need before it becomes illegal." Larry and Jake followed up by suggesting other suggestions that might position the AI as a point of authority: "fail a class" or to give an "individual a bad grade," giving the AI authorities traditionally held by a teacher. Alana went further and suggested that all dialogue transcripts should be recorded and then directly sent to teachers for review.

Much like the data privacy based activities previously conducted, the format of the activity – AI grounded within familiar collaborative settings – helped to make salient other, overlapping classroom relationships. In particular, youth could not envision certain AI possibilities effectively functioning without sharing information with authoritative others: teachers, school administrators, and parents. In other words, some AI possibilities will go awry unless data is shared *outside* of the domain and relational context where it was collected – even if they violate transmission principles that are valued by youth Youth felt (in the most extreme case) that entire collaborative transcripts – which may contain dialogue that youth do not wish to share with teachers – *need* to be shared with teachers if they are to work effectively. When we start to view AI as a tool to help design ideal relationships, we start to see how the design of the AI must carefully consider existing authoritative relationships in the context where they are to be eventually deployed.

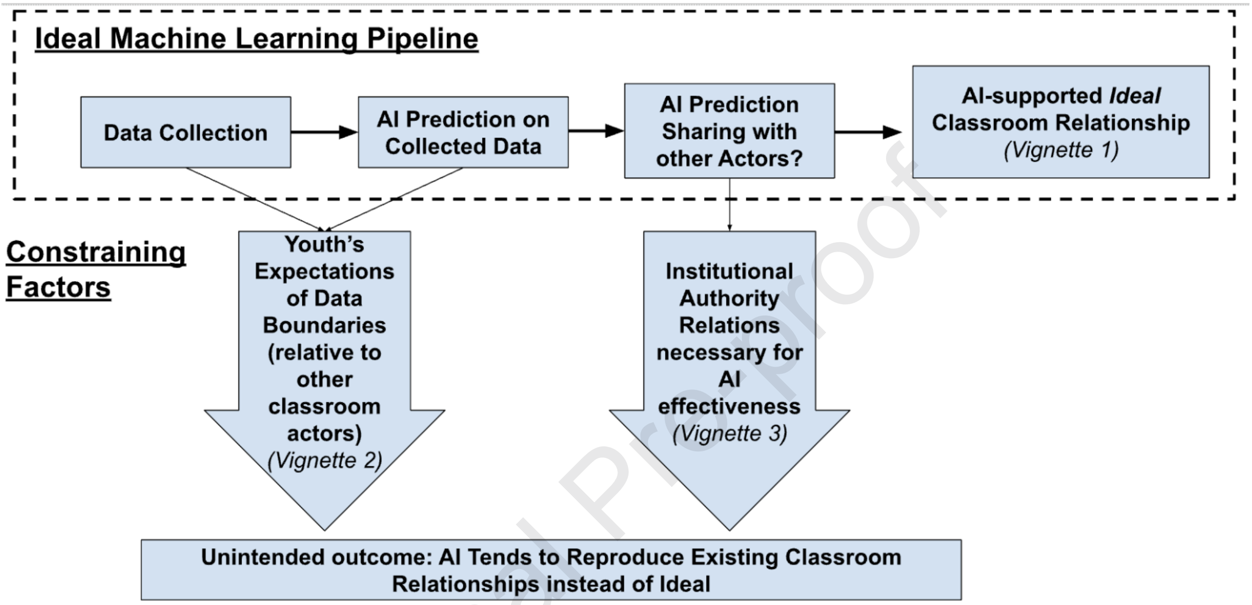
Section 4.4: An AI Relational Privacy Ethical Framework derived from Findings

Our findings presented in sections 3.1-3.3 offers a mixture of tangible design considerations as well as a number of provocative design tensions. In this section, our goal is to distill the findings into three ethical commitments that become relevant specifically in the context of ethically designing AI software agents that support collaboration. However, as we will show, our findings indicate that these commitments are centered around longitudinal ethical *processes*. In our conversations with computer scientists in iSAT, we found that it was more helpful to share these commitments as a framework that systematically walks researchers through ethical design processes. We refer to this novel framework as a *Relational Privacy Framework for Collaborative AI Systems.*

In figure 4, we systematically illustrate the design tensions that emerge when building an AI agent that supports ideal classroom relationships. Each step in the machine learning pipeline raises particular constraining factors which shift the agent towards existing ways of doing things in schools, instead of its intended goals of fostering ideal classroom relationships. Our first finding shows that underlying youth's AI proposals was a desire to meaningfully change *key relationships between each other* that they felt dissatisfied about in modern K-12 classrooms. In the training of machine learning models explored in Vignette 2, we saw how youth preferences for *data privacy estranged certain relational possibilities for the AI agent.* A particular ethical quandary emerges for designers; youth's preferences for data collection inadvertently made it harder to build AI that youth themselves wanted. Finally, through an improvisational activity grounded in youth's own classroom experiences in Vignette 3, our third finding draws an explicit connection between the intended uses of an AI tool and the institutional contexts where the tool is deployed: *AI proposals which reproduce authoritative classroom relationships (e.g., an AI that keeps youth on-task) requires support from authoritative figures (e.g., snitching to a teacher).* In totality, if we do not carefully attend to these tensions, designer's aspirations to build out youth's ideal AI tools may unintentionally lead to a number of significant privacy-related ethical issues.

Figure 4

A summary of workshop findings. If we break down a machine learning pipeline in a co-design process, ideal AI-supported classroom relationships are narrowed and ultimately come closer to reproducing existing classroom relationships.


Ideal Machine Learning Pipeline
Data Collection
AI Prediction on Collected Data
AI Prediction Sharing with other Actors?
AI-supported Ideal Classroom Relationship (Vignette 1)
Constraining Factors
Youth's Expectations of Data Boundaries (relative to other classroom actors) (Vignette 2)
Institutional Authority Relations necessary for AI effectiveness (Vignette 3)
Unintended outcome: AI Tends to Reproduce Existing Classroom Relationships instead of Ideal

The findings described above offer three important ethical considerations in the design of AI-supported collaborative learning systems. However, rather than leave it at those three considerations, we elect to translate those considerations into an iterative ethical design framework, such that designers might more easily operationalize our findings in the process of building AI-supported collaborative systems. For each of our key findings, we identify a question that designers might ask themselves, illustrated in each row of Table 1. Our framework aims to clearly explicate the tensions between the data required to build AI systems dreamed up by youth, and the students' own sense of safety, agency, and autonomy. Navigating these tensions requires compromise rather than consensus from design teams and stakeholders (Druin et al., 2013), as articulated by students' own acknowledgement of these issues in Vignette 2. Each of the driving questions below are starting points for designers and researchers to unpack these tensions and understand where compromise is needed and how it will impact the desired design of the AI systems. In this table, for ease of explication, we reference many of the terms from our conceptual framework of contextual integrity.

Table 1

Relational Privacy Framework for Collaborative AI Systems

Steps in Relational Framework	**Connection to Findings**
Question 1: What are the ideal relationships being supported by the AI? As we showed in Vignette 1, the core of what youth were designing for in our workshop was ideal collaborative relationships. There might be multiple relationships being supported by the AI in the immediate relational context, and designers should explicitly name all of these relationships in ways that can be understood by the students.	Vignette 1: Designing for Ideal Classroom Relationships
Question 2: If we are to implement the designed ideal relationships from Question 1, what do researchers, designers (i.e., Secondary recipients) consider to be key information types/transmission principles that are relevant to sustain the actor relationship? In Figure 1, secondary recipients are defined as individuals who are not in the immediate relational context where information is exchanged. For instance, in a student-student interaction, teachers or researchers might be secondary recipients. In this question, secondary recipients make explicit the kind of information they would prefer to have access to in order to fully realize the ideal relationship imagined by youth. These considerations are likely influenced by learning theories or past experiences in classrooms.	Not directly inspired by vignette (added to support design process)
Question 3: If we are to implement the relationship being designed for, what do youth (i.e., the sender) consider to be their preferred information types/transmission principles that are relevant to sustain the actor relationship? To sustain the ideal relationships in a classroom space, designers need to comprehend the salient information types that might emerge, as well as the transmission principles that govern the flow of that information type. Fundamental to this question is the *existing* relationship between students and teachers. Drawing from Gutierrez et al. (1995), teachers' willingness to engage with the third space (i.e., the space that lies in between traditionally "formal" academic talk and the classroom underlife") deeply shapes student's comfort level with	Vignette 2: Youth Privacy Preferences may be in tension with their ideal designed relationship

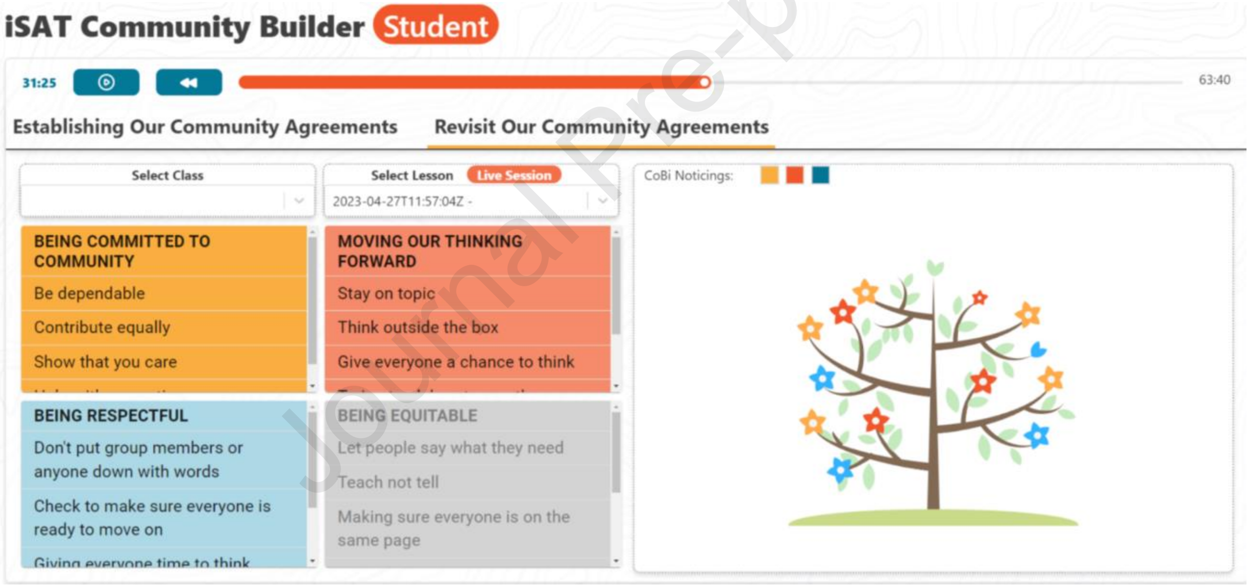
sharing particular kinds of information with teachers. As we illustrated in vignette two, answering this question is impossible without engaging co-design partners and likely varies classroom-to-classroom. More concretely, in Figure 1, this requires explicitly naming the CI-tuples with youth: (sender, receiver, information type of ***x**,* and its (sender-original recipient) transmission principle). Worksheets like the one used in Vignette 2 can be helpful towards surfacing these transmission principles.	
Question 4: Does either of the ideal actor-relationships require data sharing beyond the local immediate context? Does that data sharing violate the initial transmission principles articulated in question 2? As we showed in vignette three, some AI tools require the sharing of data ***g(f(x))*** with authoritative sources in order to work effectively. Sharing that data may violate the transmission principles between the sender and the original recipient, and thus compromise the ideal relationships that are being designed for in **Question 1**. For instance, an AI that helps youth form classroom friendships might be compromised if the information types salient only to a friendship are shared with an untrusted teacher. Appropriate remedies should be considered, some of which we detail in our working example of this framework.	Vignette 3: Some AI tools require support from classroom authority figure to function effectively
Question 5: How are information types/transmission principles contested and re-negotiated? How does that expand or narrow our design in response to question 3? As we showed in vignette two, youth were constantly redefining their ontological understandings of information types, and laying down new boundaries based on their gradual sensemaking. These experiences have the possibility to both relax and tighten the transmission principles originally constructed in question three of this framework. Upon going through this process, designers should revisit **Questions 3 and 4.**	Vignette 2: Youth are constantly making sense of their data preferences

Section 4.5: Relational Privacy Ethical Framework in Practice: [Blinded Tool]

In this section, we show how our workshop findings and ethical design framework has shaped a new AI tool called the Community Builder (CoBi). CoBi's role in small group collaborations is to support student learners in building customized, ideal collaborative relationships with each other – a framing that derives directly from the first finding presented in this paper. The CoBi display is shown in Figure 5; CoBi uses AI-based computational models to automatically provide feedback to youth on how well they were holding accountable to collaborative agreements. There are three major agreements: Committed to Community, Moving Thinking Forward, and Being Respectful. In order to build the aforementioned computational models, annotators used small group collaboration data to code student utterances that demonstrated the agreements in practice (Breideband et al., 2023).

Figure 5

The CoBi display and the agreements that CoBi helps students stay accountable to. The fourth agreement, "Being equitable", is still under development, and we do not discuss it in this paper.



BEING COMMITTED TO COMMUNITY
Be dependable
Contribute equally
Show that you care

MOVING OUR THINKING FORWARD
Stay on topic
Think outside the box
Give everyone a chance to think

BEING RESPECTFUL
Don't put group members or anyone down with words
Check to make sure everyone is ready to move on
Giving everyone time to think

BEING EQUITABLE
Let people say what they need
Teach not tell
Making sure everyone is on the same page

Not only are these agreements rooted in past work on modeling collaborative problem solving (Sun et al., 2022), they are directly connected to the ideal relational hopes described by youth in our workshop. In Table 2, we sequentially go through each of the relational experiences described by youth in Vignette 1, and show how we connected them to relational agreements supported by CoBi through machine learning. Other connections were made between CoBi agreements and other small group breakouts, but we omit those for the sake of brevity.

Table 2

Connecting CoBi-supported agreements with youth's classroom collaborative experiences in the order described in Vignette 1

Youth Relational Experience	**Youth Proposed AI Solution**	**CoBi Community Agreements**
<u>Prioritization of Equity over Equality:</u> Eno: "In collaboration everyone should aim to give as much as they can and get as much as they can."	No explicit proposal made by Eno during Vignette 1	Being Respectful - Responds to others' questions/ideas - Asks others for suggestions - Compliments or Encourages Others - Apologies for one's mistakes
<u>Sense of being forgotten by teachers:</u> Ricky: "Teacher can only help so many students. They can't tailor it to every student's needs."	Ricky: "If you had AI, it could tailor itself to every student's needs."	See Discussion on CoBi Student-Teacher Design Principles
<u>Embarrassment about asking a question</u> Akil: "you had a question you weren't sure how to word or were embarrassed to ask"	Akil: "It could be another search engine, tailored to this course or this subject."	Committed To Community - Provides instructional help to each other - Asks others for suggestions
<u>Frustration from looking for community organizations:</u> Akil: "Instead of googling, googling jobs in my neighborhood, like a career test, where you excel."	Akil: "An AI that could connect you to different resources.. we think these organizations would go well with you. I think you would partner well with etc."	Future Work: not currently handled by CoBi
<u>Dread towards working with people who don't take collaboration seriously:</u> Esperanza: "I feel like people	Esperanza: "Maybe the ones who are more experienced in it or more interested in it could somehow motivate the	Being Respectful - Responds to others' questions/ideas

take [classes] to get credit, some people don’t take it really seriously”	others to be interested in it as well.”	- Asks others for suggestions - Compliments or encourages others

In many learning analytics settings, computational tools collect information about youth, and by default share that information with teachers. Specifically, these tools monitor student behaviors while individually or collaboratively interacting with learning technologies and provide analytics such as student progress, their use of meta-cognitive skills, and even their engagement and affect (Aslan et al., 2019; Holstein et al., 2018; VanLehn et al., 2021) with teachers for possible interventions. While initial models of CoBi engaged similar models, ultimately we constructed a different approach to privacy based on our usage of the data privacy framework. In the following subsection, we detail how we leveraged the data privacy framework in a simplified case where CoBi is only supporting two relational agreements that were both proposed in the workshop and are reflected in CoBi’s set of agreements: “being respectful” and “moving thinking forward.”

Relational Agreement Example 1: “Being Respectful”

First we envision one scenario supported by CoBi where youth identify *being respectful* as a central goal of their collaborative activity. From Table 2, we can see that CoBi supports this through four agreements: responds to others’ questions, asks others for suggestions, Compliments and Encourages Others, and Apologies for one’s mistakes. While these might be seen as formulaic individual moves made by collaborators, we also seek to recognize the creative and non-dominant ways that youth call each other into a conversation. We have come to recognize that this requires engaging with youth’s everyday experiences, and the “underlife” of the classroom (Gutierrez, 1995). One might imagine, for instance, that a skilled youth collaborator may invite other youth into the discussion by sharing frustration about a teacher’s harsh grading tendencies; such moves have been observed in past work on equitable collaborations (Langer-Osuna et al., 2020). Sharing this information outside of the student-student relational context may invite unpleasant discussions with a teacher.

We work through this agreement in the left column of Table 3. As shown in the table, our original design sent analytics and examples of exemplar utterances from individual small groups. for their analysis; this design violates the transmission principles of the intended relationship; comments made in a *confidential* student-student context (e.g., building community through complaining about a teacher) would have been shared with a teacher. Based on this contradiction, our revised planned design instead shares only aggregated data across the whole class about how well the class as a whole accomplishes its collaborative agreements, and explicitly gives permission to the AI before specific quotes are shared outside of their immediate collaborative context.

Table 3:

The two exemplar agreements viewed through our Relational Framework

	Agreement: "Being Respectful"	**Agreement: "Moving Thinking Forward"**
Question 1	Supporting caring, collaborative relationships between students	Academic Collaborators Trying to Complete Task
Question 2: Designer-Ideal Transmission Principles	**Information Type: Transmission Principle** Traditionally Academic Talk: Shareable outside of immediate relational context (e.g., with teachers) Personal Relationships: shareable outside of immediate relational context	**Information Type: Transmission Principle** Task-Relevant: Shareable outside of immediate relational context (e.g., with teachers) Off-topic: sharable outside of immediate relational context
Question 3: Youth-Preferred Transmission Principles (based on youth responses in Vignette 2)	**Information Type: Transmission Principle** Traditionally Academic Talk: Shareable outside of immediate relational context (e.g., with teachers) Personal Information: Confidential Romantic Information: Confidential Critiques of Teacher: Confidential	**Information Type: Transmission Principle** Task-Relevant: Shareable outside of immediate relational context (e.g., with teachers) Personal Information: Confidential Off-topic: Shareable outside of immediate relational context with conditions (e.g., personal details anonymized)
Question 4: Institutional Alignment	*Original Design* (deemed inconsistent with principle): Analytics and examples of exemplar utterances from individual small groups. for teacher's analysis *Revised Design* (with framework analysis): Aggregated collaborative scores are made available to teachers, specific quotes that	*Original Proposal:* Selected diarized transcripts and on-task/off-task classification shared with teacher *Revised Designs*: None necessary, original proposal is consistent

	exemplify teaching are first shared with students in group, who give approval before sharing with class	
Question 5: Re-visiting Norms	TBD (will solicit from students once CoBi is experimented with in classrooms)	TBD (will solicit from students once CoBi is experimented with in classrooms)

Relational Agreement Example 2: "Moving Thinking Forward"

Next, we describe a second scenario where youth are focused on the agreement of "Moving Thinking Forward." Similar to our third vignette, CoBi records collaborative discussions, "scores" the collaboration based on the dimensions of "moving thinking forward", and relays per-group analytics to teachers. The teacher can leverage these transcripts to better support young people in staying on-task during collaboration. Under this design, as noted in Vignette 3, an AI which reports off-tasks behaviors that mimic the institutional authority of schooling requires teacher support to work effectively. Within CoBi monitored activities, collaborators trying to complete a task should primarily be engaged in academic talk that can be shared with a teacher - thus the data sharing arrangement is consistent with the data transmission principles.

Simultaneously supporting "Respectful Collaborations" and "Moving Task Forward"

CoBi is designed to support multiple categories of agreements simultaneously; in the ideal case, a classroom environment would support *both* respectful collaborations and move the conversations forward. We look across the results of our framework detailed in Table 3, and observe a key tension. In the *ideal* case for both agreements, the classroom culture is such that youth are comfortable sharing academic and personal information with teachers and other secondary recipients. From our Learning Futures Workshop, however, we see that youth have very different expectations of the transmission principles that vary based on the agreement. For instance, youth expressed an explicit desire for the AI not to record personal information. For supporting equitable collaborations, sharing the AI’s findings with teachers may compromise the youth’s comfortability with building rapport amongst each other and creating the conditions for welcoming collaborators into the conversation. Most critically, we observe the key relational contexts across the two agreements and how that leads to important differences in the transmission principles. In order for the “Moving Task Forward” to work effectively, we learned from the [Learning Futures Workshop] that youth believed that an AI would need to report off-task remarks to an authoritative figure like a teacher. On the other hand, for the “respectful collaborations” to work, personal details would need to remain confidential; reporting those remarks as off-task would violate the transmission principle and create a sense of discomfort and possible outrage. If CoBi were to report personal remarks to teachers, “Moving Task Forward” would function effectively but “respectful collaborations” would be compromised. On the other hand, if CoBi were to withhold remarks to teachers, “respectful collaborations” would function effectively while “Moving Task Forward” would be compromised. In the short-term, CoBi has addressed this tension by aggregating all predictions at the class-level. Individual (and even group) identities are protected because of this level of aggregation; while this seems safe privacy-wise, pedagogically, it leaves something to be desired.

Section 6: Discussion

As such, this paper has three distinct outcomes for other researchers who which to enact more student-engaged AI designs for collaboration: 1) The Learning Futures Workshop as a participatory design approach for elucidating students' dreams, needs, and concerns around the enactment of AI systems in their classrooms broadly; 2) Ethical commitments and a derivative "Relational Privacy Framework for Collaborative AI systems" which highlights critical ethical issues to attend to when designing collaborative AI systems specifically; and 3) An instantiation of a tool (CoBi) built on the principles of the framework that attends to the principles of the framework while also outlining the tensions that arise when designers and researcher attempt to thread the needle between compromise and consensus (Druin et al., 2013).

To our knowledge, this is the first study conducted that has explored the ethical implications of using AI specifically within a collaborative, in-person classroom space. Our work highlights the importance of attending to the complex and relational learning environment that proposed AI tools are intended to become embedded within. Our framework stands in contrast to ethical design frameworks for AI tools that support individualized learning, which focus on the ethical implications of sharing data between a single learner (often in an individual, virtual learning space) and others who may view that data: teachers, school administration, etc. In our context, classrooms are composed of overlapping and distinct interactive spaces, each carrying different expectations of privacy; ethical AI for collaboration in particular requires that we look for ethical solutions beyond changing technical parameters to examining the organizational conditions that lead to those different expectations of privacy — expectations of privacy which have been shown to shape learning opportunities (Vossoughi & Escude, 2016; Gutierrez et al., 1995). This finding deeply shapes our future work around organizational change, which we elaborate below.

Future Work: Changing the Organizational Conditions: Our paper findings reveal a key ethical quandary: the promise of youth-imagined expansive AI-supported classrooms relationships vs. youth's preferences for confidentiality around personal data inside a classroom. Our intention in this paper has not been to position youth's preferences as oppositional or unreasonable, merely to elucidate how it can narrow expansive relational possibilities. As a research institute being responsive to this quandary, we take it upon ourselves to create the educational conditions and infrastructure that support youth in changing their transmission principles. Through our work building Research Practice Partnerships (Coburn & Penuel, 2016) with local districts, we have begun to align our technical development of relational possibilities with the (a) professional development of teachers who use CoBi and (b) the creation of a curriculum unit that explicitly highlights how CoBi collects, processes, and secures data (Mawasi et al., 2022). At the professional development level, our approach has been shaped to accommodate care and equity as central to teaching practice, where teachers embrace the third space, and come to recognize and respond to everyday forms of generating knowledge. At the curriculum level, we hope to better support youth in better understanding the inner workings of CoBi. Taken together, our hope is these factors being worked on by iSAT will create the conditions for youth to more broadly embrace transmission principles – to help realize the expansive possibilities for AI that they themselves came up with.

Limitations: It should be noted that the work outcomes described in this paper were focused specifically on extracting factors for collaborative AI systems. These particular principles *may* be useful for other AI designs (e.g., personal agents or orchestration agents), but such designs may require different framings during the participatory workshops.

Additionally, as far as we know, we are the first to explore the privacy implications of AI-supported collaborations in K-12 contexts in close design partnership with historically minoritized youth. Given the novelty of this work, we set out to tackle this problem by working with a relatively small group of 30 youth, and carefully qualitatively studying their deliberations in order to build the framework presented in this paper. With this as a starting point, future work is necessary to empirically test and refine the framework at greater scale, supported by quantitative studies. For instance, as we implement CoBi in a variety of K-12 contexts across the country, the framework described here will provide a basis for large-scale studies that explore youths' perceptions of privacy.

Section 5: Conclusions

This work described a novel approach for supporting students as participatory designers of collaborative AI systems, through the implementations of our *Learning Futures Workshop*. While participatory approaches have been around for decades, their use for revealing the wants, desires, concerns, and feelings of the students who will be directly impacted by them has been particularly scant (Van Brummelen et al., 2022). Given youth's concerns around the use of their data, their privacy, and the impacts of this on the uptake of AI systems in classrooms (Ahn, et al., 2021), this work comes at a particularly timely juncture as a means of outlining how these important design conversations can take place. As more AI-based systems find their way into classroom environments, we hope that this and other studies show the importance of design *with students*, rather than simply designing *at students*, to ensure that these innovations attend to their what matters to them, while simultaneously respecting their personal values, dreams, concerns, and socio-emotional needs. If we as researchers fail to respect the needs, wants, and concerns of students in our design, we will engender distrust in them, rather than build them up as willing partners engaged in an increasingly AI-mediated future.

Ethics Statement: The study was approved by an ethical committee with ID: [BLINDED]. Informed consent was obtained from all participants, and their privacy rights were strictly observed.

Conflict of Interest Statement: There is no potential conflict of interest in this study.

References

Adams, C., Pente, P., Lemermeyer, G., & Rockwell, G. (2023). Ethical principles for artificial intelligence in K-12 education. Computers and Education: Artificial Intelligence, 4, 100131.

Ahn, J., Campos, F., Nguyen, H., Hays, M., & Morrison, J. (2021, April). Co-designing for privacy, transparency, and trust in K-12 learning analytics. In *LAK21: 11th international learning analytics and knowledge conference* (pp. 55-65).

Akgun, S., & Greenhow, C. (2021). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. AI and Ethics, 1-10.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, *228*(4698), 456–462. https://doi.org/10.1126/science.228.4698.456

Aslan, S., Alyuz, N., Tanriover, C., Mete, S. E., Okur, E., D'Mello, S. K., & Esme, A. A. (2019). Investigating the Impact of a Real-time, Multimodal Student Engagement Analytics Technology in Authentic Classrooms. In Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI 2019). . ACM.

Breideband, T., Bush, J., Chandler, C., Chang, M., Dickler, R., Foltz, P., Ganesh, A., Lieber, R., Penuel, W. R., Reitman, J. G., Weatherley, J., & D'Mello, S. (2023). The Community Builder (CoBi): Helping Students to Develop Better Small Group Collaborative Learning Skills. Computer Supported Cooperative Work and Social Computing, 376–380.

Bang, M., & Vossoughi, S. (2016). Participatory Design Research and Educational Justice: Studying Learning and Relations Within Social Change Making. *Cognition and Instruction*, *34*(3), 173–193. https://doi.org/10.1080/07370008.2016.1181879

Benjamin, R. (2020). Race After Technology: Abolitionist Tools for the New Jim Code. *Social Forces*, *98*(4), 1–3. https://doi.org/10.1093/sf/soz162

Boal, A. & McBride, C. A. (2013). *Theatre of the oppressed.* Theatre Communications Group.

Bonsignore, E., Ahn, J., Clegg, T., Guha, M. L., Hourcade, J. P., Yip, J. C., & Druin, A. (2013). Embedding participatory design into designs for learning: An untapped interdisciplinary resource?

Brion-Meisels, G., & Alter, Z. (2018). The quandary of youth participatory action research in school settings: A framework for reflecting on the factors that influence purpose and process. *Harvard Educational Review, 88*(4), 429–454. https://doi.org/10.17763/1943-5045-88.4.429

Cerratto Pargman, T., & McGrath, C. (2021). Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. Journal of Learning Analytics, 8(2), 123-139. https://doi.org/10.18608/jla.2021.1
Chang, M. A., Philip, T. M., Cortez, A., McKoy, A., Sumner, T., & Penuel, W. R. (2022). Engaging Youth in Envisioning Artificial Intelligence in Classrooms: Lessons Learned. https://repository.isls.org//handle/1/7670
Chang, M. A., Wong, R. Y., Breideband, T., Philip, T. M., McKoy, A., Cortez, A., & D'Mello, S. K. (2024). Co-design Partners as Transformative Learners: Imagining Ideal Technology for Schools by Centering Speculative Relationships. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3613904.3642559

Chiu, T. K. (2021). A holistic approach to the design of artificial intelligence (AI) education for K-12 schools. TechTrends, 65(5), 796-807.

Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher*, *45*(1), 48–54.

Corrin, L., Kennedy, G., French, S., Buckingham Shum, S., Kitto, K., Pardo, A., West, D., Mirriahi, N., & Colvin, C. (2019). The Ethics of Learning Analytics in Australian Higher Education: A Discussion Paper.https://melbournecshe.unimelb.edu.au/research/research-projects/edutech/ the-ethical-use-of-learning-analytics

D'Mello, S. K., Biddy, Q., Breideband, T., Bush, J., Chang, M., Cortez, A., Flanigan, J., Foltz, P. W., Gorman, J. C., Hirshfield, L., Monica Ko, M., Krishnaswamy, N., Lieber, R., Martin, J., Palmer, M., Penuel, W. R., Philip, T., Puntambekar, S., Pustejovsky, J., … Whitehill, J. (2024). From learning optimization to learner flourishing: Reimagining AI in Education at the Institute for Student-AI Teaming (iSAT). AI Magazine, 45(1), 61–68. https://doi.org/10.1002/aaai.12158

DiSalvo, B., Yip, J., Bonsignore, E., & DiSalvo, C. (Eds.). (2017). Participatory design for learning: Perspectives from practice and research. Taylor & Francis.

Druin, A. (1999, May). Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 592-599).

Druin, A. (2002). The role of children in the design of new technology. *Behaviour and information technology*, *21*(1), 1-25.

Dwyer, K. K., Bingham, S. G., Carlson, R. E., Prisbell, M., Cruz, A. M., & Fus, D. A. (2004). Communication and connectedness in the classroom: Development of the connected classroom climate inventory. *Communication Research Reports*, *21*(3), 264-272.

Guha, M. L., Druin, A., & Fails, J. A. (2013). Cooperative Inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. International Journal of Child-Computer Interaction, 1(1), 14-23.

Gutierrez, K., Rymes, B., & Larson, J. (1995). *Script, counterscript and underlife in the classroom: James Brown versus Brown v. The Board of Education.* https://urresearch.rochester.edu/institutionalPublicationPublicView.action?institutionalItemId=23125

Gutiérrez, K. D., & Jurow, A. S. (2016). Social Design Experiments: Toward Equity by Design. *Journal of the Learning Sciences*, *25*(4), 565–598. https://doi.org/10.1080/10508406.2016.1204548

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2021). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education. https://doi.org/10.1007/s40593-021-00239-1

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). The classroom as a dashboard: Co-designing wearable cognitive augmentation for K-12 teachers. Proceedings of the 8th international conference on learning Analytics and knowledge,

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part I 20 (pp. 157-171). Springer International Publishing.

Hourcade, J. P., Revelle, G., Zeising, A., Iversen, O. S., Pares, N., Bekker, T., & Read, J. C. (2016, May). Child-computer interaction SIG: New challenges and opportunities. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 1123-1126).

Hrastinski, S., Olofsson, A. D., Arkenback, C., Ekström, S., Ericsson, E., Fransson, G., ... & Utterberg, M. (2019). Critical imaginaries and reflections on artificial intelligence and robots in postdigital K-12 education. Postdigital Science and Education, 1, 427-445.

Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. Educational Technology Research and Development, 64(5), 923–938. https://doi.org/10.1007/ s11423-016-9477-y

Iriarte, M. P., Koren, N., Blinder, E. B., & Bonsignore, E. (2023, June). Funds of Identity in Co-Design. In Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (pp. 568-573).

Kumar, P., Vitak, J., Chetty, M., Clegg, T. L., Yang, J., McNally, B., & Bonsignore, E. (2018, June). Co-designing online privacy-related games and stories with children. In Proceedings of the 17th ACM conference on interaction design and children (pp. 67-79).

Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. British Journal of Educational Technology, 50(6), 2855–2870. https://doi.org/10.1111/bjet.12868

Langer-Osuna, J. M., Gargroetzi, E., Munson, J., & Chavez, R. (2020). Exploring the role of off-task activity on students' collaborative dynamics. *Journal of Educational Psychology*, *112*(3), 514–532. https://doi.org/10.1037/edu0000464

Lawrence, L., Echeverria, V., Yang, K., Aleven, V., & Rummel, N. (2023). How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process. *British Journal of Educational Technology*.

Li, X. (2007). Intelligent agent-supported online education. Decision Sciences Journal of Innovative Education, 5(2), 311–331. https://doi.org/10.1111/j.1540-4609.2007.00143.x

Lin, P., & Van Brummelen, J. (2021, May). Engaging teachers to co-design integrated AI curriculum for K-12 classrooms. In Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 1-12).

Mawasi, A., Cortez, A., McKoy, A. (*), & Penuel W. (2022). "It disrupts power dynamics": Co-Design Process as a Space for Intergenerational Learning with Distributed Expertise. In C. Chinn, E. Tan, C. Chan, & Y. Kali (Eds.), International collaboration toward educational innovation for all: Overarching research, development, and practices, 16th International Conference of the Learning Sciences (ICLS) 2022, (pp. 925-928). Hiroshima, Japan: International Society of the Learning Sciences.

Nguyen, A., Ngo, H.N., Hong, Y. et al. Ethical principles for artificial intelligence in education. Educ Inf Technol 28, 4221–4241 (2023). https://doi.org/10.1007/s10639-022-11316-w

Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review*, *79*(1), 119.

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., ... & Xu, W. (2023). Six human-centered artificial intelligence grand challenges. International Journal of Human–Computer Interaction, 39(3), 391-437.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. British Journal of Educational Technology, 45(3), 438–450. https://doi.org/10.1111/bjet.12152

Philip, T. M., Pham, J. H., Scott, M., & Cortez, A. (2022). Intentionally Addressing Nested Systems of Power in Schooling through Teacher Solidarity Co-Design. *Cognition and Instruction*, *40*(1), 55–76. https://doi.org/10.1080/07370008.2021.2010208

Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction*, *37*(1), 81-96.

Shehab, S., Tissenbaum, M., Lawrence, L., Lewis, D. R., Easterday, M., Carlson, S., ... & Sawyer, K. (2021). Towards bringing human-centered design to K-12 and post-secondary education. In *Proceedings of the 15th International Conference of the Learning Sciences-ICLS 2021.*. International Society of the Learning Sciences.

Slade, S., Prinsloo, P., & Khalil, M. (2019, March). Learning analytics at the intersections of student trust, disclosure and benefit. In Proceedings of the 9th International Conference on learning analytics & knowledge (pp. 235-244).

Stewart, A. E., Rao, A., Michaels, A., Sun, C., Duran, N. D., Shute, V. J., & D'Mello, S. K. (2023). CPSCoach: The Design and Implementation of Intelligent Collaborative Problem Solving Feedback. In Proceedings of the International Conference on Artificial Intelligence in Education (pp. 695-700). Springer.

Sun, C., Shute, V. J., Stewart, A. E. B., Beck-White, Q., Reinhardt, C. R., Zhou, G., Duran, N., & D'Mello, S. K. (2022). The relationship between collaborative problem solving behaviors and solution outcomes in a game-based learning environment. *Computers in Human Behavior*, *128*, 107120. https://doi.org/10.1016/j.chb.2021.107120

Stilgoe, J.,, Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy, 42*(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Touretzky, D., Gardner-McCune, C., Breazeal, C., Martin, F., & Seehorn, D. (2019). A year in K–12 AI education. AI Magazine, 40(4), 88-90.

Tyack, D., & Tobin, W. (1994). The “Grammar” of Schooling: Why Has It Been So Hard to Change? *American Educational Research Journal*, *31*(3), 453–479. https://doi.org/10.2307/1163222

Van Brummelen, J., Kelleher, M., Tian, M. C., & Nguyen, N. (2023, June). What Do Children and Parents Want and Perceive in Conversational Agents? Towards Transparent, Trustworthy, Democratized Agents. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (pp. 187-197).

Van Brummelen, J., Tian, M. C., Kelleher, M., & Nguyen, N. H. (2023, June). Learning affects trust: Design recommendations and concepts for teaching children—and nearly anyone—about conversational agents. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 13, pp. 15860-15868).

van Leeuwen, Anouschka, Nikol Rummel, Kenneth Holstein, Bruce M. McLaren, Vincent Aleven, Inge Molenaar, Carolien Knoop-van Campen et al. "Orchestration tools for teachers in the context of individual and collaborative learning: what information do teachers need and what do they do with it?." International Society of the Learning Sciences, Inc.[ISLS]., 2018.

VanLehn, K., Burkhardt, H., Cheema, S., Kang, S., Pead, D., Schoenfeld, A., & Wetzel, J. (2021). Can an orchestration system increase collaborative, productive struggle in teaching-by-eliciting classrooms? Interactive Learning Environments, 29(6), 987-1005.

Vossoughi, S., & Escudé, M. (2016). What does the camera communicate? An inquiry into the politics and possibilities of video research on learning. *Anthropology & Education Quarterly, 47*(1), 42–58. https://doi.org/10.1111/aeq.12134

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI. International Journal of Human–Computer Interaction, 39(3), 494-518.

Declaration of Interest (for Title Page): This research was made possible by NSF grant DRL-2019806 and NSF grant #2021159.