Designing Stakeholder-Based Pedagogy for AI Ethics Education: Insights from a Multi-Institutional Case Study

Yifan Zhang, Harini Suresh, Julia Netter
Department of Computer Science, Brown University
yifan_zhang4@alumni.brown.edu, {harini_suresh, julia_netter}@brown.edu

Abstract

As AI systems increasingly shape real-world decisions, there is growing urgency to teach students not only how to build these systems but also how to consider the social impact of their design choices. One promising approach explored in computing education is the stakeholder-based pedagogy, which, in AI courses, encourages students to examine how AI development affects different groups and to design with those needs and values in mind. In this pilot study, we conducted interviews with 14 instructors across three U.S. institutions to understand how they design and implement stakeholder-based modules in undergraduate AI ethics education. Our findings reveal two core pedagogical strategies: empathy-driven role play and structured stakeholder analysis, as well as cross-cutting challenges, including the heavy labor of curating materials, limited integration of stakeholder reasoning into the broader AI curriculum, time constraints, difficulty embedding stakeholder reasoning into technical instruction, and unclear assessment methods. We also examine how these challenges are shaped by the curricular structures and institutional contexts in which the modules were implemented. Drawing on these insights, we offer recommendations to strengthen stakeholder-based AI ethics education at both the micro level (e.g., establishing shared resource repositories and assessment frameworks across institutions) and the macro level (e.g., prioritizing stakeholder reasoning as a core AI competency and fostering channels for interdisciplinary and academia-industry dialogue). This work contributes to ongoing conversations in AI ethics education on how to better prepare students to become socially responsible AI practitioners through meaningful stakeholder analysis and engagement.

Introduction

As AI technologies become increasingly integrated into diverse industries and aspects of human life, there is growing recognition within the computing education field of the need to teach AI across varied pedagogical contexts, from K-12 (Chiu 2021; Grover 2024) to higher education (Allen, McGough, and Devlin 2021; Niousha et al. 2025) settings. Traditionally, AI courses in higher education have focused on cultivating students' understanding of the technical aspects of AI (Saltz et al. 2019).

However, as AI systems continue to demonstrate concerns such as biased outputs, privacy and consent violations in data collection, and representational and participatory inequities, both in high-stakes domains like healthcare (Norori et al. 2021; Cross, Choma, and Onofrey 2024) and criminal justice (Wang et al. 2024), and in general-purpose technologies like large language models (Bender et al. 2021; Heidt 2024), scholars have called for the integration of ethics into undergraduate AI curricula (Borenstein and Howard 2021). The goal is to prepare students, many of whom will become future developers of these systems, to recognize and address the societal and ethical challenges associated with AI.

In response, AI ethics education has grown as a field over the past decade, shaped by broader movements within computing education that advocate for embedding ethical reasoning throughout the curriculum. Recent analyses of global AI curricula suggest that ethics is often taught in disciplinary silos, focusing separately on building trustworthy systems ("build"), evaluating ethical tradeoffs ("assess"), or protecting those impacted by technical solutions ("govern"), without sufficiently integrating these approaches (Javed et al. 2022). This fragmentation risks leaving students without the tools to understand how ethical concepts translate into technical implementation, or how such implementation must be situated within broader social and political contexts and informed by engagement with affected communities. This challenge is also faced by AI practitioners when developing ethical AI systems (Pant et al. 2024).

To address this gap, one promising approach that has been explored in prior work is what we refer to a *stakeholder-based pedagogy*. According to the National Institute of Standards and Technology (NIST), a stakeholder is any individual or organization that has an interest in a system or its characteristics (NIST 2022). This pedagogy emphasizes that ethical reasoning in AI must account for how system design choices affect diverse groups, particularly those historically excluded from or harmed by technological systems, and aims to equip students to recognize, analyze, and meaningfully engage with the values and perspectives of these groups throughout the development process. Importantly, stakeholder engagement throughout the AI development lifecycle is also described as a crucial component of responsible AI work in the 2019 EU Ethics Guidelines for Trustworthy AI and in recent research on RAI practices in the real world (European Commission 2019; Bach et al. 2025).

While there are varied conceptualizations of stakeholder-based pedagogy, they follow shared focuses and strategies. This approach acknowledges the role of diverse stakeholders in technology development and emphasizes the importance of integrating their perspectives into the teaching of computing ethics. Through specific instructional strategies, students are encouraged to consider the needs and perspectives of stakeholders (e.g., users, community members, regulatory bodies) in the context of particular technologies. They may synthesize this understanding through outputs such as discussions, written responses, and technical implementations. This approach can support the teaching of AI ethics by grounding ethical issues in real-world scenarios and encouraging students to explore how understanding and addressing the needs of diverse stakeholders can help tackle these ethical challenges and contribute to systems that better reflect needs of affected communities.

Despite its promise, the experiences of instructors implementing stakeholder-based pedagogy in practice remain understudied. Our work examines how instructors apply stakeholder-based approaches in undergraduate AI classes and the challenges they encounter. We focus on two main research questions:

- **RQ1:** How do instructors implement stakeholder-based pedagogy in undergraduate AI courses, and what beliefs, attitudes, and positionalities shape their approach?
- **RQ2:** What challenges do instructors encounter when implementing this pedagogy, and how do they envision improving their implementation?

To address these questions, we present a qualitative pilot study of stakeholder-based pedagogy in undergraduate AI ethics education. Drawing on semi-structured interviews with 14 instructors across three U.S. institutions, our study offers a grounded perspective on how instructors design and implement stakeholder-focused modules and assignments within AI courses. This cross-institutional lens provides initial insights into shared and distinct challenges, instructional strategies, and contextual constraints that shape the integration of stakeholder reasoning into technical AI curricula. Building on these insights, we offer recommendations for researchers, educators, industry, policy, and community partners to strengthen stakeholder-based AI ethics education at both the micro level and the macro level. Our work contributes to ongoing conversations in AI ethics education and responsible AI by offering an instructor-centered account of stakeholder pedagogy and outlining paths to enhance this approach. In doing so, we call for AI ethics education that not only fosters students' ability to reason about ethics abstractly but also prepares them to design AI systems through hands-on work informed by an understanding of the complex social, institutional, and industry contexts and actors with which such systems interact. Equipped with these skills, students will be better prepared to develop AI responsibly in their professional careers.

Related Work

Evolving Approaches to AI Ethics Education

AI ethics is defined by (IBM 2024) as "a multidisciplinary field that studies how to optimize the beneficial impacts of AI while reducing risks and adverse outcomes." Inspired by calls to include AI ethics as part of AI literacy (Knoth et al. 2024) and competency, and to integrate it into the AI curriculum, many organizations have developed AI competency frameworks that incorporate principles such as transparency, safety, non-discrimination, and ethics by design (UNESCO 2024), as well as dedicated AI ethics competency frameworks for higher education (Bruneault, Sabourin Laflamme, and Mondoux 2022). Building on these frameworks, institutions have implemented AI ethics education with the predominant goal of teaching students these principles (Aler Tubella, Mora-Cantallops, and Nieves 2024).

In the past few years, AI ethics education has highlighted a growing trend toward real-world, socially grounded pedagogy. An earlier study shows that many instructors have started to incorporate news articles and current events to help students engage with the concrete impacts of AI systems, although some instructors still embed ethical discussions within technical content in a way that is limited to "technical" interpretations of concepts such as fairness or privacy, without sufficient explanation of the broader social, historical, or political contexts in which these concepts operate (Garrett, Beard, and Fiesler 2020). More recently, instructors have been employing domain-specific case studies to contextualize guidelines on AI ethics in areas such as art creation, business, and biomechanics (Deb et al. 2025) and encouraging students to analyze ethical dilemmas encountered in real-world AI applications (Usher and Barak 2024).

Building on these content-driven approaches, instructors have also been adopting more hands-on and collaborative formats for teaching AI ethics. A recent systematic literature review by (Wiese et al. 2025) illustrates that hands-on activities have become the most common instructional strategy. Increasingly, these activities are grounded in real-world scenarios, including team-based audits of algorithmic systems (Stoyanovich 2022) and exercises such as fine-tuning or jail-breaking AI models based on concrete AI ethics and policy principles (Weichert et al. 2025). Discussion-based formats are the second most frequent strategy, supporting collaborative reasoning and preparing students for responsible public discourse around AI—a merit also echoed in another literature review by (Ranade and Saravia 2024) and in broader trends in computing ethics instruction (Brown et al. 2024). For instance, students might read case studies about autonomous vehicles and collaboratively analyze their interests as stakeholders, as well as assessing the system's potential risks (Hishiyama and Shao 2022).

Collectively, these approaches reflect a pedagogical shift in which instructors are increasingly teaching AI ethics not only as a set of abstract principles or technical constraints, but as practices embedded in social and political realities that require interdisciplinary engagement. This shift raises questions about what specific pedagogical frameworks can support students in connecting ethical reasoning to concrete design choices affecting specific communities. Our work explores stakeholder-based pedagogy as one such framework.

Stakeholders as a Pedagogical Tool in AI and Computing Ethics

A small but growing body of work explores the use of stakeholders as a teaching framework in computing ethics education. Across six representative studies (Ren and Fisler 2023; Netter et al. 2025; Shen et al. 2021; Horton et al. 2022; Castro et al. 2023; Silva, Castro, and Guimaraes 2021) spanning introduction to computing (Horton et al. 2022; Ren and Fisler 2023), AI (Shen et al. 2021; Castro et al. 2023), software engineering (Silva, Castro, and Guimaraes 2021), and computer systems courses (Netter et al. 2025), instructors introduced stakeholder-based modules that encouraged students to grapple with conflicting perspectives and make value-based decisions. These modules took various forms, from scenario-based discussions (Horton et al. 2022) to project-based learning with external stakeholders (Silva, Castro, and Guimaraes 2021), and aimed to cultivate ethical reasoning, empathy, and stakeholder analysis skills.

Although promising, existing implementations face several limitations. Most are one-off activities that do not include direct interaction with real stakeholders. Only half (Shen et al. 2021; Netter et al. 2025; Silva, Castro, and Guimaraes 2021) involved technical planning and implementation; the rest relied on written reflections or theoretical analysis. In addition, there is disagreement about whether such modules should promote specific ethical values (e.g., privacy) or focus on helping students reason through competing stakeholder needs without prioritizing particular outcomes (Horton et al. 2022; Netter et al. 2025). These tensions mirror broader debates in computing ethics education about normative framing versus analytical neutrality.

It is important to note that few of these studies are situated within AI-specific contexts. Of the six reviewed, only two (Shen et al. 2021; Castro et al. 2023) focus on scenarios grounded in AI or machine learning. Our study addresses this gap by adopting an instructor-centered, interview-based approach across multiple universities. Through semi-structured interviews, we examine how instructors design, deliver, and evaluate stakeholder-based modules in undergraduate AI courses, and the institutional and pedagogical constraints they navigate. This grounded account builds on prior findings while providing new insight into the factors that shape stakeholder-based pedagogy in AI classrooms, as well as the successes and challenges instructors encounter in practice.

Methods

We conducted semi-structured interviews with 14 instructors from 10 undergraduate-level AI courses at three private U.S. universities, each with an undergraduate enrollment under 8,000 students. These institutions were selected because they have established computing ethics curricula integrated across a range of computing courses, including AI. None of the three institutions requires a standalone ethics course for completion of the computer science major; instead, ethics content is embedded within technical courses across the curriculum.

ID	Role	Disc.	Involvement
P1	UG TA	CS	Led discussion (B)
P2	UG TA	Ling.	Led discussion (B)
P3	Faculty	CS	Reviewed discussion materials (B)
P4	UG TA	CS	Led discussion (B)
P5	UG TA	CS	Led discussion (B)
P6	Faculty	CS	Co-developed content (B)
P7	UG TA	CS	Led discussion (B)
P8	UG TA	CS	Created homework questions (B)
P9	UG TA	CS	Led discussion (B)
P10	Postdoc	CS/Phil	Led modules (A)
P11	Postdoc	Phil.	Led modules (A)
P12	Postdoc	Phil.	Led modules (A)
P13	Postdoc	Phil.	Led modules (A)
P14	Grad TA	CS	Created lab questions (C)

Table 1: Participant Background and Teaching Involvement

The courses taught by our participants cover topics such as general AI, machine learning, deep learning, computer vision, reinforcement learning, and computational linguistics. While open to non-CS majors, these courses typically require prerequisites in data structures, algorithms, and relevant mathematical foundations. Most were large-format courses enrolling between 50 and over 100 students.

The structure of stakeholder-based ethics instruction varied across institutions. At Institution A, ethics content was delivered as a single one-hour class session facilitated by postdoctoral teaching fellows. Institution B incorporated three to six structured discussion sections or homework assignments per semester, led by dedicated undergraduate teaching assistants (primarily CS majors) who curated ethics content. At Institution C, ethics material was introduced through written or discussion-based prompts in labs and homework, designed by student teaching assistants. Across all three contexts, stakeholder-based pedagogy formed part of a broader institutional effort to integrate ethical reasoning into technical AI instruction.

We recruited participants through targeted email outreach to instructors who had taught relevant AI courses during or before Fall 2024, after receiving IRB approval from our institution. We sent interested instructors a consent form outlining study goals and procedures and they provided signed consent prior to participating in a one-hour online interview with the lead author. Interviewees included faculty members, postdoctoral instructors, and undergraduate or graduate teaching assistants involved in module design and facilitation.

Table 1 summarizes participants' instructional roles, disciplinary backgrounds, and teaching involvement across the three institutions. Here, UG means undergraduate, Grad means graduate, CS means computer science, Phil. means philosophy, Ling. means linguistics. Institution labels are put in parenthesis (A, B, or C). All TAs and postdocs who led modules were also involved in content creation.

Informed by our specific research questions, we designed open-ended interview questions to probe how instructors approached the design, implementation, and evaluation of stakeholder-based modules or assignments. Sample questions included:

- "How did you determine the specific content of your stakeholder module or assignment? What sources or guidelines did you consult?"
- "How did you determine the format of the stakeholder module or assignment?"
- "What feedback or reactions from students did you look for to gauge their understanding or the success of the module or assignment?"
- "What would you hope to improve if you were to teach the module or implement the assignment again? What forms of support would be helpful?"

This broader scope allowed us to understand the context in which stakeholder-based pedagogy was deployed and to probe for stage-specific details, instructor attitudes, challenges, and opportunities for future improvements. The full list of interview questions can be found in this supplementary document. [1]

All interviews were recorded, transcribed, and analyzed using deductive thematic analysis, with initial codes developed directly from our research questions and interview protocol. We created 11 initial codes: four capturing logistical details (description of the module, design strategies, instructional goals, and collaboration during the design process), two capturing instructor attitudes and positionality, and five capturing perceived successes, challenges (organized by stage of the instructional workflow), and envisioned improvements. The lead author conducted the coding in Google Sheets, with the other two authors reviewing coded segments and providing feedback for iterative refinement. A total of 137 quotes were coded. Following review and consolidation, the codes were grouped into seven categories most relevant to our research questions. From these seven codes, we synthesized our findings into six overarching themes, which are elaborated in the Findings section. See the codes below:

- **Description of module or assignment** (the structure and format of stakeholder-based module or assignment)
- **Instructor positionality and attitudes** (beliefs, disciplinary backgrounds, and teaching philosophies)
- **Collaboration** (experiences with interdisciplinary or intra-institutional coordination; closely related to the code above)
- **Challenges in curation** (difficulty sourcing materials and designing the module or assignment)
- **Challenges in implementation** (logistical or pedagogical barriers during in-class delivery)
- **Challenges in assessment** (how student engagement and success were defined and evaluated)
- **Planned improvements** (reflections on desired changes and forms of institutional or structural support)

[1]https://docs.google.com/document/d/1dE7XBAwSfC7wk1kVpHFWZs_ZKMImszJw1Xt9sEFmAic/edit?usp=sharing

Findings

Here, we present key insights from the interviews, organized around six recurring themes that address our research questions. The first two themes illustrate how instructors implement stakeholder-based pedagogy in undergraduate AI courses and beliefs, attitudes, and positionalities that shape their approach and experiences (**RQ1**), while the remaining four themes highlight the challenges they encounter in the process and how they envision future improvements of their implementations (**RQ2**). Together, these themes capture instructors' strategies, constraints, and differing perspectives on the goals and practices of this pedagogical approach.

Modes of Engagement: Balancing Empathy and Ethical Analysis

Instructors employed two primary approaches to engage students in stakeholder-based thinking around AI technologies. One approach emphasized empathy and perspective-taking through interactive role-based discussions, while the other focused on structured stakeholder and ethical analysis that students would conduct individually. Although both aimed to cultivate critical thinking and awareness of how design decisions affect stakeholders, each approach reflected distinct pedagogical philosophies and was shaped by the institution's instructional model.

Empathy-driven approaches were commonly associated with Institution B, where stakeholder-based pedagogy was implemented through periodic small-group discussion sections. These sessions, typically aligned with technical topics taught in class, were facilitated by undergraduate teaching assistants specializing in ethics. In these sessions, students were assigned stakeholder roles and asked to advocate for those perspectives in real-world scenarios. These exercises were typically followed by discussions where students debated design decisions or responsibility in AI incidents, while staying in character. One instructor noted that this method helped students "embody a perspective other than their own" (P2) and "wear the shoes of those people [stakeholders of an AI system]." (P5) Instructors observed that students were more engaged "when they were playing a role" instead of answering as themselves (P2), especially when engaging with peers in small groups. Several instructors also emphasized how this approach helped students empathize with often-overlooked populations, such as artists affected by copyright violations (P7), while benefiting from peer perspectives in discussion (P9).

In comparison, analysis-oriented approaches were more commonly found at Institution A, where stakeholder-based modules were embedded as one-time, one-hour ethics sessions delivered to the entire class by postdoctoral instructors. This model emphasized structured reasoning and often drew from the instructors' disciplinary backgrounds in philosophy. Specifically, instructors used structured stakeholder analysis, ethical frameworks, and sometimes technical components to guide learning. They asked students to identify relevant actors, map harms and benefits, and analyze cases as third-party observers. In some assignments, students also acted as developers to design or implement algorithmic solutions. One instructor described a module where students examined the effects of Uber's algorithm on stakeholders such as companies, drivers, and customers. After coming up with relevant stakeholders and analyzing each of their values, students "wrote their own algorithm" and justified their design decisions based on stakeholder impact (P10). Instructors using this approach aimed to help students understand how design choices affect "different kinds of actors" and how ethical reasoning can inform concrete design decisions (P12). While elements of this approach also appeared at the other two institutions, they were typically integrated into homework and lab questions.

Although instructors did not express a desire to entirely switch approaches, many acknowledged trade-offs within their own practices. Those who employed empathy-driven methods noted the need for more scaffolding and conceptual depth. For instance, one instructor described stakeholder role-play discussions as occasionally feeling "a bit tokenistic." (P1) Another reflected that while role-playing was "a good starting point," it didn't always push students to synthesize multiple perspectives or fully grasp underlying, systemic tensions within the scenario (P2). Others pointed to gaps between students' lived experience and the stakeholder roles they were asked to inhabit. As one instructor explained, "The population here [at the university] is very, very different than...truck drivers and such [whose employment could be affected by AI]...There's only so much we can do [to help students understand these stakeholders' values]." (P6)

In comparison, instructors who prioritized structured analysis acknowledged that heavy emphasis on formal ethical theory could make the material feel abstract or inaccessible. Some expressed interest in more interactive formats that enabled collaborative reflection. For example, one instructor noted that introducing welfare theory into a module on Uber "maybe wasn't necessary," and suggested replacing it with small-group design critiques to help students more intuitively grasp stakeholder values (P10).

Across both approaches, we observed that instructors' choices were closely shaped by institutional format and pedagogical role. Empathy-based methods flourished in smaller, discussion-driven sections, while analysis-oriented strategies aligned with one-off lectures or structured assignments.

Disciplinary Boundaries and Instructor Positionality

Instructors' disciplinary backgrounds and personal experiences shaped how they designed and delivered stakeholder-based content. While most participants were situated within computer science departments, especially at Institutions B and C, many drew on training or interests in other fields, such as linguistics, philosophy, literature, or law. These interdisciplinary perspectives influenced the case studies they selected, the stakeholders they emphasized, and the frameworks they used to guide student learning.

At Institutions B and C, where ethics instruction was often facilitated by undergraduate or graduate TAs, instructors frequently drew on personal interests or coursework to shape the content of assignments and discussion sections. For example, one instructor used her background in linguistics to incorporate speech communities as stakeholders in discussions about machine translation systems for low-resource languages (P2). Another instructor designed a case study on stakeholder perspectives in an educational AI development scenario, incorporating considerations of data privacy laws and regulations (P8). These domain-specific connections were established with the aim of making the content feel more concrete and personally relevant to students.

In contrast, instructors at Institution A, primarily postdocs with backgrounds in philosophy, leaned more towards introducing formal ethical frameworks such as social philosophy, welfare theory, or decision theory. These frameworks were used to scaffold student analysis by offering structured tools for identifying stakeholders, articulating values, and reasoning through tradeoffs in system design (P10, P11, P12). While these approaches added conceptual rigor, they were also shaped by the institutional format: short, lecture-based modules intended to deliver structured ethical reasoning within a limited time frame.

Still, some instructors expressed ambivalence about integrating disciplinary content from outside computer science. One instructor noted that some of their colleagues were hesitant to introduce humanities or social science frameworks, fearing that they might oversimplify complex fields or lack the time to meaningfully engage students with them (P2).

Instructors also reflected on their own authority in the classroom and how it influenced the delivery of stakeholder-based content. Whether or not they embraced interdisciplinary materials, many described a commitment to maintaining neutrality. One TA emphasized the importance of avoiding value imposition, noting, "When you are in a position of power, it's very easy for students to be like, the teacher feels a certain way about something, I should feel that way too." (P4) A postdoctoral instructor echoed this sentiment, highlighting the disciplinary ethos of philosophy: "Philosophers are not in the business of giving answers… they're going to teach skills for reasoning about ethics, but they're not going to teach answers to questions." (P11)

Designing Stakeholder-Based Content: The Labor of Curation

While instructors actively embraced stakeholder-based pedagogy, many described the process of identifying relevant case studies, selecting meaningful stakeholder perspectives, and designing the most appropriate delivery formats as time-consuming, resource-intensive, and improvisational, often lacking standardized guidelines or institutional support.

Many instructors noted that curating stakeholder configurations that felt both balanced and thought-provoking was a significant challenge. "As far as actually coming up with what stakeholders we showed to the students," one explained, "we wanted it to be diverse, like each week to be slightly different." (P13) Another instructor emphasized that the goal of stakeholder curation was to introduce meaningful tension into the scenario: "Not every stakeholder has the same goals in mind—sometimes they're completely opposite, sometimes they're not, but just because of one difference it could lead to very different results." (P4)

These challenges were further intensified by the fast-paced nature of AI as a field. "The field is moving so fast... there's these new examples and these new things (AI technologies) that are going on in new ways," one instructor observed (P9). To meet these expectations, instructors described spending substantial time searching for case studies and designing modules. Some estimated spending "five to ten hours," or even "a few weeks," to develop each module, not including preparation or grading (P13, P14).

In addition to the time burden, instructors frequently voiced frustration with the lack of shared resources or curricular guidance. When choosing case studies and stakeholders, many relied on personal interests, news articles, or book chapters. One instructor described the process as "very vibe-based" and lacking a clearly defined method (P13). These burdens were especially acute for instructors at Institution B, who often lacked formal training in ethics or stakeholder theory. In contrast, instructors at Institution A found it easier to select stakeholder-based case studies and design modules because they could draw on their prior academic training in philosophy and ethics. This background, along with the support of a network of postdoctoral colleagues with similar expertise, helped them identify meaningful frameworks and stakeholder configurations for their modules. Within Institution B, concerns were shared across roles; for example, one faculty member without a background in ethics expressed hope that "people who know more about this stuff than I do can help [throughout the process]." (P9)

Others also highlighted the broader uncertainty around how to curate materials on stakeholder engagement practices. As one instructor remarked, "That's one of the big questions we need to do a better job addressing—what stakeholder engagement actually looks like," noting that even in industry, it remains "not really standardized, so it might be even more difficult to teach students how to do that." (P2)

Without clear guidelines, shared curricular resources, or institutional support, the work of designing stakeholder-based modules remained highly individualized and uneven, placing greater demands on instructors without formal ethics backgrounds, who had to navigate decisions about materials, structure, and depth largely on their own.

Time Constraints and Curricular Fit

Another common concern raised by instructors was the challenge of time constraints, particularly for approaches grounded in structured stakeholder analysis. While empathy-based, role-driven activities could often be integrated more quickly into discussion sections, analytical frameworks were described as requiring not only time for context and perspective-building but also additional space to introduce structured methods of analysis. These elements demanded deeper scaffolding and extended classroom time to ensure meaningful student engagement.

This issue was often exacerbated when stakeholder-focused modules were not well integrated into the broader course or departmental curriculum, or when instructors lacked power over how much instructional time was allocated to these modules, such as at Institutions A and C. One instructor from Institution A described feeling like they had "one chance to tell these people to be more thoughtful about the work they're doing." (P12) Another instructor proposed a two-part model, "one to set up the problem, and then one to talk through the different ways of dealing with it", but was only allotted a single one-hour session to deliver the module (P12). In another setting, an instructor noted that "each week we had maybe 10 minutes" to devote to ethics-related content in lab sessions, making in-depth stakeholder-based questions difficult to implement in practice (P13). By contrast, this challenge was less pronounced at Institution B, where instructors, despite being student teaching assistants, were granted autonomy over the design and timing of stakeholder modules. This flexibility allowed them to more effectively scaffold activities across the semester and adapt the pacing to fit course needs.

At a broader curricular level, instructors also pointed to the lack of scaffolding across the AI curriculum. They emphasized that the complexity of stakeholder analysis often requires prior knowledge, such as identifying stakeholder groups, articulating values, or conducting trade-off analyses, that students may not yet have encountered. As one instructor from Institution C put it, "There's just no time to introduce that concept [of stakeholders]... if the students don't have a common set of knowledge to draw on." (P12) Taken together, these reflections suggest that instructors often face structural and curricular barriers that limit the depth, continuity, and overall impact of stakeholder-based modules and assignments.

Embedding Stakeholder Reasoning into AI System Design

Several instructors across the institutions emphasized the importance of incorporating design or implementation-oriented components into stakeholder-based modules. Instructors who advocated for this approach argued that integrating stakeholder analysis into teaching the technical design of AI systems, rather than treating it as a separate or theoretical activity, could help students better grasp how their decision decisions affect different groups in concrete and often unequal ways.

One instructor described their ideal scenario as one in which "technical components [are] incorporated into the assignments so that people can see more specifically how... what feel like very tangible [design] decisions can actually really affect the behavior of the system in a way that could have different effects on different groups," such as in the context of natural language processing (P3). As they put it, if students are only taught to "implement algorithms" and leave it to others to "do this critical thinking and decide how to deploy them [the algorithms they built]," they may come to view stakeholder engagement as irrelevant to their role, an attitude she described as one that "absolves you of the responsibility to think about it." (P3)

For some instructors, this approach also helped reframe stakeholder discussions as opportunities for agency and creativity, rather than moral paralysis. Reflecting on a discussion section he led, one instructor explained that "it highlighted how these systems [i.e., NLP for low resource languages] can be used for good" when stakeholders are engaged responsibly (P1). Others echoed that without giving students concrete strategies for addressing stakeholder needs, the session could feel overwhelming to students and "in some ways, demoralizing when you're like, I'm impacting all of these people. And then you have no idea where to go from there." (P12)

Instructors shared several strategies they have used to embed stakeholder reasoning into design-oriented activities. In a section on low-resource language translation, students read technical papers, analyzed design choices, and then reflected on how speech communities could be included as active participants in the process (P1, P2). Another discussion section in a computer vision course asked students to design new human-AI interaction workflows for medical diagnosis and justify their decisions in relation to different stakeholder groups such as doctors, patients, and hospitals. As the instructor who led this section explained, "I think having people designing a new system was a really good way to get people active about this." (P4)

Looking ahead, one instructor suggested that if students have already been building stakeholder analysis and value reasoning skills before taking upper-level AI courses, they could be asked to take on more advanced tasks with less scaffolding, such as "come up with their own application, come up with a reward function, and then to do this type of [stakeholder] analysis." (P13)

Ultimately, instructors saw teaching stakeholder reasoning not as separate from AI system design, but as a necessary approach to help students understand their agency and responsibility as future developers. In this framing, stakeholder analysis becomes less about abstract discussions of stakeholder values or hypothetical trade-offs detached from implementation, and more about embedded decision-making within the design process itself.

Assessing Learning Outcomes: What Counts as Success?

While instructors utilized various strategies to gauge whether students were engaging meaningfully with stakeholder-based content, many described the assessment of learning outcomes as an ongoing challenge, especially in large, fast-paced courses where reflections and discussions were frequent but difficult to track at scale. Furthermore, without a shared definition of what counts as success, instructors were often left to navigate these questions on their own, without the time, resources, or clarity to do so with confidence.

Some instructors described looking for signs of thinking-outside-the-box in students as an indicator of success. One noted that a particularly strong student response in a stakeholder analysis activity about Uber's algorithm involved a student discussing how the company's business model could shape the algorithm's impact on drivers, something that "wasn't already included in my framing of the question." (P10) This suggested that the student was drawing on additional knowledge and considering complex, interconnected factors that influence decision-making. Another echoed that they were "looking for students' ability to kind of expand their thinking," such as by identifying additional stakeholder groups beyond those introduced in the prompt, especially those less directly impacted but still part of "the bigger picture." (P12, P5)

However, many instructors acknowledged that there was no clear consensus on what success in stakeholder-based modules should look like, or how it should be assessed. This uncertainty shaped the ways instructors approached evaluation, many of which were effort-based instead of with clear assessment rubrics. As one instructor put it, "It's hard to come up with a rubric [for these questions]" and assess student skills given the open-ended and reflective nature of the assignments (P8). This lack of a standard assessment structure reflected a deeper philosophical uncertainty around the goals of stakeholder-based pedagogy itself. As one instructor explained, "People disagree a lot on whether we are just teaching good reasoning or if we're supposed to be actually imparting values or getting people to have the substantively right views... it's challenging to assess either of those things." (P11)

Large class sizes, which are common in AI classes, further increased difficulty for assessment. As one instructor shared, "We didn't really have the time to do a comprehensive read through [of student reflections after the discussion sections]," because with approximately 200 students in the class, each attending two discussion sections and submitting two reflections per session, the course generated around 800 responses for just two instructors to manage (P1). Given this volume, instructors were often limited in how deeply they could assess students' growth in stakeholder analysis and reasoning skills, especially without a standardized assessment framework.

Discussion

Our findings reinforce prior literature showing that stakeholder-based pedagogy can broaden students' perspectives and cultivate ethical reasoning (Shen et al. 2021; Castro et al. 2023), while also revealing persistent challenges, both familiar (e.g., risk of superficial engagement, difficulty connecting students to stakeholder whose experiences are significantly different from their own) (Ren and Fisler 2023; Castro et al. 2023) and unexplored (e.g., rapid pace of technological change in the AI field, lack of consistent industry models for stakeholder engagement, uncertainty over what constitutes "good" stakeholder analysis). We also find that institutional structures and instructor positionality influence how stakeholder-based modules are designed and delivered, shaping both their integration into the curriculum and the resources available for implementation.

Furthermore, although the term responsible AI did not explicitly appear in our interviews or instructional materials, many of the pedagogical choices described by instructors (e.g., case studies chosen, technical methods taught) align closely with responsible AI principles, such as ensuring the fairness and inclusiveness of AI applications and protecting the privacy of data subjects (Microsoft 2024). This suggests that our findings could inform efforts to bridge responsible AI work and educational practice from a stakeholder analysis and engagement perspective.

In the following discussion, we draw on these insights to identify concrete directions for future research and pedagogical support, as well as broader structural and cultural shifts needed to more effectively implement stakeholder-based pedagogy in the AI curriculum.

Cross-Sector Collaboration and Pedagogical Support

Many instructors, particularly those without formal training in ethics or stakeholder theory, described designing stakeholder-based modules as time-intensive, highly individualized, and lacking clear guidance. To reduce this burden, we recommend developing shared pedagogical resources, such as adaptable case studies (e.g., from the AI Incident Database (McGregor 2020) and participatory AI research involving specific communities (Queerinai et al. 2023; Lee et al. 2019; Suresh et al. 2022; Zhang et al. 2023)), stakeholder analysis templates, and discussion prompts tailored to different AI subfields. These resources could allow instructors to focus on pedagogy and student engagement rather than content creation.

Our findings also underscore the value of interdisciplinary perspectives in shaping stakeholder-based content. Instructors often drew on backgrounds in fields such as linguistics, law, and literature to inform case framing and stakeholder identification. This supports prior calls to foster deeper collaboration between ethics and technical instructors (Raji, Scheuerman, and Amironesei 2021). Institutions could formalize these partnerships through co-designed modules, team-teaching models, or cross-departmental collaborations that enable instructors to leverage disciplinary strengths while addressing concerns about misrepresenting unfamiliar frameworks.

Crucially, such collaborations must be equitable. Instructors from non-CS disciplines should be given sufficient authority and space to meaningfully shape module content, rather than being limited to surface-level contributions. As our findings from Institution A suggest, one-time guest-led ethics sessions may constrain instructors' ability to meet their pedagogical goals or integrate meaningfully with the broader course design.

Besides, partnerships with responsible AI and human-computer interaction researchers, as well as industry and policy organizations, could help develop grounded and sustainable models for teaching stakeholder engagement. Such collaborations could integrate clarified terminology on stakeholder involvement (Kallina, Bohné, and Singh 2025), delineate the scope of such involvement (Ajmani, Abdelkadir, and Chancellor 2025), and draw from established stakeholder-centered methods (e.g., participatory design (Delgado et al. 2023), value-sensitive design (Zhu et al. 2018), user-centered design (Liao et al. 2024)). These partnerships could offer students insight into how stakeholder engagement is, or is not operationalized in real-world AI development and help instructors teach both frameworks and the practical tensions that emerge in applied contexts.

To support these cross-sector and interdisciplinary efforts, we advocate for formal structures such as collaborative teaching fellowships, interdisciplinary working groups, and dedicated conference tracks that recognize and incentivize the labor of stakeholder-based ethics instruction in AI education.

Curricular Scaffolding and Thorough Integration

Many instructors emphasized that teaching stakeholder reasoning effectively requires more than a well-designed module; it requires embedding those modules into a broader curricular sequence. When ethics is introduced only in upper-level AI courses, instructors must compress both foundational concepts and application-specific reasoning into a single class session, limiting depth and consistency.

To address this issue, stakeholder-based thinking should be scaffolded across the CS curriculum. Introductory courses could introduce key ideas such as identifying stakeholders, understanding how design decisions create harms and benefits for different groups, and analyzing tradeoffs between competing stakeholder needs. Mid and upper-level AI courses could then build on this foundation by analyzing domain-specific tensions in areas like NLP, computer vision, or machine learning. By the time students reach these courses, ideally they would be prepared to engage in more complex activities such as technical implementation or analyzing systems in multi-stakeholder contexts.

Importantly, for this curricular scaffolding to succeed, instructors and departments must explicitly prioritize stakeholder understanding, analysis and engagement as core competencies that students need to develop through AI (and more generally, computing) courses, rather than treating them as secondary or optional topics. Supporting this goal, future research might explore how early and repeated exposure to stakeholder concepts affects students' ethical reasoning in AI courses.

Clearer Assessment Strategies

Another major challenge instructors reported was assessing student learning outcomes in stakeholder-based modules. Without a shared definition of what successful stakeholder reasoning looks like, instructors, particularly those teaching large classes, were often left to evaluate student performance based on individual interpretation. This lack of standardization can make it difficult to gauge effectiveness, ensure consistency across sections, or iterate on pedagogy over time.

To address this gap, we recommend developing clearer and adaptable assessment frameworks, including rubrics tailored to the specific learning goals of each module or assignment. These frameworks could be informed by collaboration with industry, which is increasingly defining the competencies required of responsible AI and AI ethics practitioners. For example, recent work on responsible AI in industry outlines roles and skills in the RAI workforce, identifying the core competencies expected of engineers, data scientists, technical researchers, and AI ethicists (Rismani and Moon 2023; Cocchiaro et al. 2025). Specifically, Rismani and Moon (2023) suggest that educators "use the list of competencies to develop a set of learning objectives and examine the efficacy of different teaching pedagogies in supporting these objectives." Building on this work, future research could investigate how stakeholder reasoning and engagement map onto these competencies and how they can be translated into actionable learning objectives.

In addition to industry alignment, assessment frameworks should be co-developed with instructors across institutions through collaborative forums or disciplinary communities. Doing so would ensure that rubrics reflect pedagogical expertise, promote shared standards for strong stakeholder reasoning, and remain adaptable to the curricular structures and constraints of different institutions. As with the design and resource development strategies discussed earlier, this effort would benefit from formal institutional support and sustained interdisciplinary collaboration.

Limitations and Future Work

All instructors in our study were affiliated with private U.S. universities, which often benefit from greater access to educational resources and established computing ethics infrastructure. This limits the generalizability of our findings to institutions with different resource levels, governance structures, or curricular priorities.

Future research should broaden both the institutional and geographic scope of studies on stakeholder-based pedagogy. Including public universities, liberal arts colleges, community colleges, non-traditional programs, and institutions outside the U.S. would help surface a wider range of pedagogical practices, constraints, and priorities. Collecting data from a larger and more diverse sample could also enable quantitative analyses to identify trends and examine how institutional factors such as available resources, labor structures, and curricular priorities shape the implementation and sustainability of stakeholder-based pedagogy.

Our study also focused exclusively on AI courses. Future work could examine how stakeholder-based pedagogy is applied in other computing subfields, such as software engineering or computer security. Comparing these implementations could help distinguish challenges unique to AI education, such as the rapid pace of technological change, from those shared across computing curricula. Such comparisons would clarify where targeted pedagogical interventions or curricular supports are most needed.

Conclusion

Our pilot study offers an instructor-centered investigation of stakeholder-based pedagogy in undergraduate AI ethics education. By drawing on interviews with 14 instructors across three U.S. institutions, we provide a grounded and comparative perspective on how stakeholder reasoning is taught in AI courses and how it is shaped by institutional structures, instructor backgrounds, and curricular constraints.

Our findings reaffirm the value of stakeholder-based pedagogy for fostering ethical reasoning and perspective-taking, particularly through strategies such as empathy-based role-playing and structured analysis. At the same time, we surface new challenges that are underexplored in prior work focusing on this pedagogy, including the difficulty of keeping pace with rapidly evolving AI technologies, the absence of clear frameworks for stakeholder engagement in industry, and the varying levels of institutional integration and instructional authority granted to ethics instructors.

In response, we propose a set of initial recommendations to support more sustainable and impactful implementations of stakeholder-based pedagogy. These include developing centralized design resources and guidelines, as well as building adaptable assessment frameworks aligned with real-world responsible AI competencies. We also call for broader cultural and structural shifts that recognize stakeholder understanding and engagement as core AI competencies, and that better support the instructional labor required to teach them by fostering structures for cross-sector collaboration and sustained interdisciplinary dialogue. These recommendations are not exhaustive, and new insights may emerge as future research expands to other types of institutions within the U.S. and internationally.

Overall, our work contributes to the growing body of research on teaching AI ethics and responsible AI by offering insights that can inform ongoing efforts within the AI ethics education community to support instructors in teaching stakeholder reasoning alongside AI system design. We hope these efforts will help better prepare students to empathize with stakeholders, engage meaningfully with their needs, and make responsible, thoughtful design decisions as future AI developers.

Acknowledgments

We are grateful to Kathi Fisler and the anonymous reviewers for their helpful feedback on different versions of this paper. This work was funded by the Brown University Socially Responsible Computing Program.

References

Ajmani, L. H.; Abdelkadir, N. A.; and Chancellor, S. 2025. Secondary Stakeholders in AI: Fighting for, Brokering, and Navigating Agency. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, 1095–1107. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714825.

Aler Tubella, A.; Mora-Cantallops, M.; and Nieves, J. C. 2024. How to teach responsible AI in Higher Education: challenges and opportunities. *Ethics and Information Technology*, 26(3).

Allen, B.; McGough, A. S.; and Devlin, M. 2021. Toward a Framework for Teaching Artificial Intelligence to a Higher Education Audience. *ACM Trans. Comput. Educ.*, 22(2).

Bach, T. A.; Kaarstad, M.; Solberg, E.; and Babic, A. 2025. Insights into suggested Responsible AI (RAI) practices in real-World settings: a systematic literature review. *AI Ethics*, 5: 3185–3232. Published online 8 Jan 2025.

Bender, E. M.; Gebru, T.; McMillan-Major, A.; and Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 610–623. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383097.

Borenstein, J.; and Howard, A. 2021. Emerging challenges in AI and the need for AI ethics education. *AI Ethics*, 1: 61–65.

Brown, N.; Xie, B.; Sarder, E.; Fiesler, C.; and Wiese, E. S. 2024. Teaching Ethics in Computing: A Systematic Literature Review of ACM Computer Science Education Publications. *ACM Transactions on Computing Education*, 24(1): Article 6, 36 pages.

Bruneault, F.; Sabourin Laflamme, A.; and Mondoux, A. 2022. AI Ethics Training in Higher Education: Competency Framework. Cégep André-Laurendeau / École des médias, UQAM.

Castro, F.; Raipura, S.; Conboy, H.; Haas, P.; Osterweil, L.; and Arroyo, I. 2023. Piloting an Interactive Ethics and Responsible Computing Learning Environment in Undergraduate CS Courses. SIGCSE 2023, 659–665. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394314.

Chiu, T. K. F. 2021. A Holistic Approach to the Design of Artificial Intelligence (AI) Education for K-12 Schools. *TechTrends*, 65: 796–807.

Cocchiaro, M. Z.; Morley, J.; Novelli, C.; Panai, E.; Tartaro, A.; and Floridi, L. 2025. Who is an AI Ethicist? An Empirical Study of Expertise, Skills, and Profiles to Build a Competency Framework. *AI Ethics*, 5: 3713–3725. Published online 12 Feb 2025.

Cross, J. L.; Choma, M. A.; and Onofrey, J. A. 2024. Bias in medical AI: Implications for clinical decision-making. *PLOS Digital Health*, 3(11): e0000651.

Deb, D.; Taylor, G.; Betz, S.; Maddux, B. A. T.; Ebert, C. E.; Richardson, F. W.; Couto, J. L. S.; Jarrett, M. S.; and Madjd-Sadjadi, Z. 2025. Enhancing University Curricula with Integrated AI Ethics Education: A Comprehensive Approach. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*, SIGCSETS 2025, 248–254. New York, NY, USA: Association for Computing Machinery. ISBN 9798400705311.

Delgado, F.; Yang, S.; Madaio, M.; and Yang, Q. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '23. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703812.

European Commission. 2019. Ethics guidelines for trustworthy AI. Accessed: 2025-08-09.

Garrett, N.; Beard, N.; and Fiesler, C. 2020. More Than "If Time Allows": The Role of Ethics in AI Education. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, AIES '20, 272–278. New York, NY, USA: Association for Computing Machinery. ISBN 9781450371100.

Grover, S. 2024. Teaching AI to K-12 Learners: Lessons, Issues, and Guidance. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, SIGCSE 2024, 422–428. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704239.

Heidt, A. 2024. Intellectual property and data privacy: the hidden risks of AI. *Nature.com.*

Hishiyama, R.; and Shao, T. 2022. Educational Effects of the Case Method in Teaching AI Ethics. In Rocha, A.; Adeli, H.; Dzemyda, G.; and Moreira, F., eds., *Information Systems and Technologies. WorldCIST 2022*, volume 468 of *Lecture Notes in Networks and Systems*, 226–236. Springer.

Horton, D.; McIlraith, S. A.; Wang, N.; Majedi, M.; McClure, E.; and Wald, B. 2022. Embedding Ethics in Computer Science Courses: Does it Work? In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1*, SIGCSE 2022, 481–487. New York, NY, USA: Association for Computing Machinery. ISBN 9781450390705.

IBM. 2024. What is AI Ethics? Accessed: 2025-08-09.

Javed, R. T.; Nasir, O.; Borit, M.; Vanhée, L.; Zea, E.; Gupta, S.; Vinuesa, R.; and Qadir, J. 2022. Get out of the BAG! Silos in AI Ethics Education: Unsupervised Topic Modeling Analysis of Global AI Curricula. *Journal of Artificial Intelligence Research*, 73: 933–965.

Kallina, E.; Bohné, T.; and Singh, J. 2025. Stakeholder Participation for Responsible AI Development: Disconnects Between Guidance and Current Practice. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, 1060–1079. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714825.

Knoth, N.; Decker, M.; Laupichler, M. C.; Pinski, M.; Buchholtz, N.; Bata, K.; and Schultz, B. 2024. Developing a holistic AI literacy assessment matrix – Bridging generic, domain-specific, and ethical competencies. *Computers and Education Open*, 6: 100177.

Lee, M. K.; Kusbit, D.; Kahng, A.; Kim, J. T.; Yuan, X.; Chan, A.; See, D.; Noothigattu, R.; Lee, S.; Psomas, A.; and Procaccia, A. D. 2019. WeBuildAI: Participatory Framework for Algorithmic Governance. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): Article 181, 35 pages.

Liao, Q. V.; Vorvoreanu, M.; Subramonyam, H.; and Wilcox, L. 2024. UX Matters: The Critical Role of UX in Responsible AI. *Interactions*, 31(4): 22–27.

McGregor, S. 2020. Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database. arXiv:2011.08512.

Microsoft. 2024. Responsible AI Principles and Approach. Accessed: 2025-08-09.

Netter, J.; Nelson, T.; Austen, S.; Lau, E.; Rusch, C.; Schwarzkopf, M.; and Fisler, K. 2025. A Stakeholder-Based Framework to Highlight Tensions When Implementing Privacy Features. *Proceedings of the USENIX Security Symposium.*

Niousha, R.; Guo, L. J.; Li, R. K.; Norouzi, N.; and Zhang, L. 2025. Comparing Artificial Intelligence Curricula in Canadian and US Universities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 29053–29061.

NIST. 2022. Systems Security Engineering, NIST Special Publication 800-160 Volume 1 Revision 1.

Norori, N.; Hu, Q.; Aellen, F. M.; Faraci, F. D.; and Tzovara, A. 2021. Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2: 100347.

Pant, A.; Hoda, R.; Spiegler, S. V.; Tantithamthavorn, C.; and Turhan, B. 2024. Ethics in the Age of AI: An Analysis of AI Practitioners' Awareness and Challenges. *ACM Trans. Softw. Eng. Methodol.*, 33(3).

Queerinai, O. O.; Ovalle, A.; Subramonian, A.; Singh, A.; Voelcker, C.; Sutherland, D. J.; Locatelli, D.; Breznik, E.; Klubicka, F.; Yuan, H.; J, H.; Zhang, H.; Shriram, J.; Lehman, K.; Soldaini, L.; Sap, M.; Deisenroth, M. P.; Pacheco, M. L.; Ryskina, M.; Mundt, M.; Agarwal, M.; Mclean, N.; Xu, P.; Pranav, A.; Korpan, R.; Ray, R.; Mathew, S.; Arora, S.; John, S.; Anand, T.; Agrawal, V.; Agnew, W.; Long, Y.; Wang, Z. J.; Talat, Z.; Ghosh, A.; Dennler, N.; Noseworthy, M.; Jha, S.; Baylor, E.; Joshi, A.; Bilenko, N. Y.; Mcnamara, A.; Gontijo-Lopes, R.; Markham, A.; Dong, E.; Kay, J.; Saraswat, M.; Vytla, N.; and Stark, L. 2023. Queer In AI: A Case Study in Community-Led Participatory AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 1882–1895. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Raji, I. D.; Scheuerman, M. K.; and Amironesei, R. 2021. You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 515–525. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383097.

Ranade, N.; and Saravia, M. 2024. Teaching AI Ethics in Technical and Professional Communication: A Systematic Review. *IEEE Transactions on Professional Communication*, 67(4): 422–436.

Ren, Y.; and Fisler, K. 2023. A Social Threat Modeling Framework to Structure Teaching about Responsible Computing. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, SIGCSE 2023, 402–408. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394314.

Rismani, S.; and Moon, A. 2023. What does it mean to be a responsible AI practitioner: An ontology of roles and skills. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, 584–595. New York, NY, USA: Association for Computing Machinery. ISBN 9798400702310.

Saltz, J.; Skirpan, M.; Fiesler, C.; Gorelick, M.; Yeh, T.; Heckman, R.; Dewar, N.; and Beard, N. 2019. Integrating Ethics within Machine Learning Courses. *ACM Transactions on Computing Education*, 19: 1–26.

Shen, H.; Deng, W. H.; Chattopadhyay, A.; Wu, Z. S.; Wang, X.; and Zhu, H. 2021. Value Cards: An Educational Toolkit for Teaching Social Impacts of Machine Learning through Deliberation. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 850–861. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383097.

Silva, L. H.; Castro, R. X.; and Guimaraes, M. C. 2021. Supporting real demands in software engineering with a four steps project-based learning approach. In *Proceedings of the 43rd International Conference on Software Engineering: Joint Track on Software Engineering Education and Training*, ICSE-JSEET '21, 50–59. IEEE Press. ISBN 9780738133201.

Stoyanovich, J. 2022. Teaching Responsible Data Science. In *Proceedings of the 1st International Workshop on Data Systems Education*, DataEd '22, 4–9. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393508.

Suresh, H.; Movva, R.; Dogan, A. L.; Bhargava, R.; Cruxen, I.; Martinez Cuba, A.; Taurino, G.; So, W.; and D'Ignazio, C. 2022. Towards Intersectional Feminist and Participatory ML: A Case Study in Supporting Feminicide Counterdata Collection. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 667–678. New York, NY, USA: Association for Computing Machinery.

UNESCO. 2024. AI Competency Framework for Students.

Usher, M.; and Barak, M. 2024. Unpacking the role of AI ethics online education for science and engineering students. *International Journal of STEM Education*, 11(35).

Wang, A.; Kapoor, S.; Barocas, S.; and Narayanan, A. 2024. Against Predictive Optimization: On the Legitimacy of Decision-making Algorithms That Optimize Predictive Accuracy. 1(1).

Weichert, J.; Dunlap, D.; Farghally, M.; and Eldardiry, H. 2025. The AI Policy Module: Developing Computer Science Student Competency in AI Ethics and Policy. arXiv:2506.15639.

Wiese, L. J.; Patil, I.; Schiff, D. S.; and Magana, A. J. 2025. AI ethics education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 8: 100405.

Zhang, A.; Boltz, A.; Lynn, J.; Wang, C.-W.; and Lee, M. K. 2023. Stakeholder-Centered AI Design: Co-Designing Worker Tools with Gig Workers through Data Probes. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394215.

Zhu, H.; Yu, B.; Halfaker, A.; and Terveen, L. 2018. Value-Sensitive Algorithm Design: Method, Case Study, and Lessons. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW).