“You Can’t Sit With Us”: Exclusionary Pedagogy in AI Ethics Education

Inioluwa Deborah Raji
Mozilla Foundation
deborah@mozillafoundation.org

Morgan Klaus Scheuerman
Information Science, University of Colorado Boulder
morgan.scheuerman@colorado.edu

Razvan Amironesei
CADE, University of San Francisco
ramironesei@usfca.edu

ABSTRACT

Given a growing concern about the lack of ethical consideration in the Artificial Intelligence (AI) field, many have begun to question how dominant approaches to the disciplinary education of computer science (CS)—and its implications for AI—has led to the current “ethics crisis”. However, we claim that the current AI ethics education space relies on a form of “exclusionary pedagogy,” where ethics is distilled for computational approaches, but there is no deeper epistemological engagement with other ways of knowing that would benefit ethical thinking or an acknowledgement of the limitations of uni-vocal computational thinking. This results in indifference, devaluation, and a lack of mutual support between CS and humanistic social science (HSS), elevating the myth of technologists as "ethical unicorns" that can do it all, though their disciplinary tools are ultimately limited. Through an analysis of computer science education literature and a review of college-level course syllabi in AI ethics, we discuss the limitations of the epistemological assumptions and hierarchies of knowledge which dictate current attempts at including ethics education in CS training and explore evidence for the practical mechanisms through which this exclusion occurs. We then propose a shift towards a substantively collaborative, holistic, and ethically generative pedagogy in AI education.

ACM Reference Format:
Inioluwa Deborah Raji, Morgan Klaus Scheuerman, and Razvan Amironesei. 2021. “You Can’t Sit With Us”: Exclusionary Pedagogy in AI Ethics Education. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT ’21), March 3–10, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3442188.3445914

1 INTRODUCTION

Over the last few years, alongside the rise of public scrutiny of the role of artificial intelligence (AI) in reifying and amplifying social inequalities, machine learning educators have begun to acknowledge the necessity of an ethics curricula in computer science programs. Khari Johnson of VentureBeat called it a “fight for the soul of machine learning” [44]. From the ACM ethics charter [2], to the call to add ethics to CS curricula [28], there has been an ongoing call to “make engineers ethical.”

However, there have also been noted observations of the failure of the approach of inserting ethics education into CS curriculum—either referenced throughout or in a standalone course—as a limited intervention for improving the outcomes of the discipline [9]. Although an important step towards informing more socially conscious system builders, it is becoming clear that proposals anchored to developing individual morality and understanding falls short of resulting in any noticeable changes to the way in which students conduct research and develop applications for deployment once they leave the classroom [35]. This is made even more evident by the consistency with which such crises continue to occur [81].

Incidents of algorithmic misuse, unethical deployments, or harmful bias cannot be addressed by developing moral integrity at an individual level. We argue that this is because the individual scope of current educational approaches neglects the fact that the current issues are more likely the result of collective failure, and more institutionalized practices accepted within the field of computer science, rather than moments of individual judgement. In fact, such challenges are inherently interdisciplinary, requiring the cooperation of stakeholders of varying expertise in business, law, and other domains in order to meaningfully address in the real world [23]. If anything, to rush CS students through heavily condensed and simplified overviews of broad ethical understanding and then position them to be the primary arbiter of change confuses the situation. This promotes the engineer’s natural inclination towards seeing themselves as a solitary saviour, to the detriment of the quality of the solution and in spite of the need for other disciplinary perspectives.

Therefore, in order to address the social impact of technical systems, including AI, we need to revisit the way we think about the norms of AI ethics education, and in particular address the tendency towards an “exclusionary” pedagogy, that further siloes CS perspectives to challenges from the necessary consideration of other approaches. This is a required first step towards the genuine interdisciplinary collaboration necessary to meaningfully address the ethical issues that continue to arise. The way in which we teach AI ethics informs the way in which practitioners are trained and reflects academic practice. Rather than exploring strategies to retrain AI scholars or practitioners—exposing them to a sprinkle of ethics and social science, and centering interventions on how to incorporate social considerations into technical expertise—we instead discuss the need to think more deeply about what it would mean to reset the pedagogy and practices of the field to shift away from this exclusionary default. Through a systematic analysis of over 100 AI ethics syllabi, we map the current situation and characterize some

FAccT ’21, March 3–10, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8309-7/21/03.
https://doi.org/10.1145/3442188.3445914

suggestions for an educational reset towards a more collaborative pedagogy, with hopefully more direct consequences on improving both industry practice and academic norms.

2 HOW COMPUTER SCIENCE PEDAGOGY LED US TO THE ETHICS "CRISIS"

We first examine the literature on how the culture of computer science has led to the current state of ethics discussions in the field, focusing specifically on computer science education research, epistemological analyses of computer science, and empirical studies of computer science classrooms and cultures. We then discuss current issues in tech ethics, primarily its insular focus on *techno-solutionism* that continues to prioritize computer science expertise and center the system itself in ethical fixes, as well as a promotion of the ideal of *ethical unicorns* or *tech saviours*, ie. technologists with shallow socio-technical understanding intent on playing the primary role in delivering complete solutions.

2.1 Historical Retrospective on CS Education Norms

In starting with computer science as a discipline, broadly, there is a heavy focus on what Eden identifies as three paradigms: technocratic, rationalist, and scientific [24]. At its simplest, the technocratic paradigm might be described as an engineering or programmatic approach, which centers the skills to build computer programs; the rationalist paradigm might be described as a mathematically theoretical approach, focusing more heavily on a priori knowledge about the underlying mathematical reality of computer systems; and the scientific paradigm might be best characterized by its focus on empiricism, seeking to more deeply understand the behaviors of computer programs. These three paradigms shape how computer science operates, through enculturation in computer science education and in industry cultures, at both the level of attitudes and values, and at the level of behaviors and practices. While there are often debates on the education of computer scientists, particularly between rationalist and technocratic approaches [20], if distilled to its most basic form, each paradigm centers what is often considered a "technical" expertise—that is, an expertise rooted in mathematics, logic, and programming.

The valuing of the technical is evident in current research on computer science education. In a survey of introductory computing education literature, Pears et al. found that computing textbooks most often focused on the correctness of syntactic structure [60], limiting what Turkle and Papert define as the "epistemological pluralism" necessary for computer scientists [74]. Turkle and Papert discuss the need to allow students as part of their education in computer science to acknowledge the possibility of interpretation, collaboration and argument as part of the practical programming experience by engaging in group-based technical project assignments.

In an analysis of computer science degree requirements at thirty-one universities in the U.S., Surakka showcased the heavily mathematical and programmatic emphasis on computer science teaching—in fact, the most "human-centered" specialization, usability, was also the least offered course as part of the analyzed degree requirements [71]. While the lack of human-centered or ethics subject matter may have changed since these studies, this snapshot in time—before a more recent push for ethics education in computer science —highlights a longstanding disciplinary norm: learning computer science has traditionally emphasized mathematical theory and engineering practices. A focus on "programming intelligence" is a well-researched cause of the high dropout rates of computer science undergraduate programs [5, 30, 36, 75]. Computer science is what Clark labels "a hard-applied discipline" [14]. Those who approach computer science problems differently are pushed out of the field, resulting in more homogeneous mindsets and practices within the discipline and unsurprising diversity deficits [49].

An imbalance between the "technical" and the "social"—the "hard" and the "soft"—which prioritizes the former, has not gone unnoticed by other scholars. In an ethnographic study of machine learning practitioners in industry, Forsythe witnessed the valuing of computer science skills, the devaluing of user needs, and the belittling of women's work (commonly characterized as social and soft) [33]. She posited that social science perspectives would improve AI by acting as a counter weight to address what she called an epistemological imbalance, though this was a view that AI researchers actively resisted at the time [31]. Later on, Wagstaff wrote that machine learning researchers, in their lack of training in understanding social contexts, often fail to create models that have real world applicability or merit; among many suggestions to remedy this failure, Wagstaff suggests the improved interaction and involvement with "the outside world" when creating models [76]. In a scathing analysis of the technocratic dominance in computer science education, Washington wrote: "With no formal courses that focus on the non-technical issues affecting marginalized groups and how to address and eradicate them, students are indirectly taught that the current status quo in computing departments and industry is not only acceptable, but also unproblematic" [77]. Tomayko equally laments that computer science education "is a story of academics struggling to fulfill industry needs with almost no support from computer science curriculum designers. It is a story of industry finally winning over some of academia to teach software engineering rather than vanilla computer science" [72]. In other words, he paints computer science education as a funnel from classroom to tech company, with little space for nuanced reflection on its foundational norms and objectives.

Since early on in the development in the field, several scholars have sounded the alarm of a need for reflection on social responsibility and meaningful interaction with other disciplines in order to make computer systems meaningful and beneficial to society [1, 37, 50, 51, 55]. However, for decades, those pleas have been largely downplayed or ignored, with one-time ethics modules or courses being rarely considered a primary focus for many students and even well intentioned attempts at integrating ethics considerations throughout CS curriculum encountering challenges with contradicting the taught paradigms of the discipline [7, 38, 67, 68]. Current approaches to ethics education appear to also burden individual faculty with the responsibility of implementing course design from scratch, even when the computer science instructors often assigned to develop these courses feel unqualified due to the limited nature of their own training [43]. Fiesler et al. further notes that, after a survey of AI ethics curricula, it is clear that the majority of ethics courses are being taught from within the discipline [29]—from computer scientists to computer scientists. A discipline which has otherwise been criticized for its lack of ethical engagement is now taking up the mantle of instilling ethical wisdom to its next generation of students.

Attitudes, values, practices, and norms all shaped by the epistemological approaches of a discipline help to explain how problems are being defined and approached, and what answers are viewed as appropriate to those problems. Certain characteristics of the pedagogical norms of computer science become limiting for the discipline's ability to address its own ethical challenges. As Clarke formulates in [14], while understanding disciplines is much more difficult and complex than ideal characterizations, it is a useful endeavor in understanding what has led to the formulation of a technocratic locus in machine learning ethics, despite the need for ethics stemming from the failure of this technocratic tendency in the first place. Like Eden, one might argue that the technocratic paradigm has dominated computer science, specifically AI, due to industry and monetary incentives driving programmatic approaches where any problem is best addressed by technologists, including the problems they create. More specifically, Agre famously characterized the nature of the conception of "reality" in AI—which posits better systems and better models as the only means of critiquing AI—as problematic in its ignorance and dismissal of sociological or critical theory [1].

Therefore, although CS, and specifically AI systems are fundamentally malleable and interpretable, computer science as a discipline remains anchored to prioritize a positivist framework for their analysis of these systems. Although students are often taught the importance of computational thinking for problem solving, they are rarely exposed to what such thinking can take away from one's ability to appropriately analyze less familiar challenges, such as this ethics crisis. There are insights that one gains from a computational lens and also insights lost. Being anchored to one perspective for so long as a discipline is a limiting factor and the historical lack of social science training in computer science has given rise to computer science researchers, teachers, and industry practitioners well-versed in techno-solutionist methodologies but not social realities, leading to systems that are—often inadvertently—inaccessible, opaque, unethical, and harmful.

2.2 Engineering Responsibility & The Individual Technologist

The notion of engineering responsibility is not new—this is Theodore Cooper's miscalculations precipitating the 1907 collapse of the Quebec bridge [61], General Motors causing thousands of avoidable deaths after neglecting to address the uncontrollable steering of its 1960 Chevrolet Corvair [54]—situations in which the harm caused is in direct consequence of a careless neglect of the details of the engineering process itself. These are the completely avoidable situations—evidence of not understanding the weight of the contribution, the impact and influence of the outcomes, and thus the need to approach things with caution. Focus on this topic of due diligence in the development of robust systems and instilling a sense of increased responsibility is fairly common—often being heavily featured in AI ethics education [52], and professional ethics codes [81].

However, the reality is that engineers are often absent or excluded from decisions that lead to harm. These decisions can often be attributed to sales people, executies, marketing and other stakeholders in a corporation—at times without even informing the technologist of the broader context of what they're working on. Similarly, problematic unethical research practices in the field are mostly enabled by a set of entrenched norms and operational structures of everything from funding to the review process and publication. Deployed CS systems are not just artifacts by themselves but a component of a complete sociotechnical system. The challenge is thus not necessarily to perfect the technical artifact, but to investigate and implement approaches to increase the participation of and communication to other stakeholders, from users to affected populations to other key decision-makers in the company to other experts studying the problem from the lens of wildly different disciplines. A step towards more inclusive design is thus less about a single engineer's efforts push to enforce their understanding of diverse representation into the worldview of the model, and more about a form of participatory design where these other stakeholders are actively and humbly welcomed to join the engineer in the creation of more just and equitable AI systems. Although blamed by the Royal Commission for the bridge's collapse, an organizational review reveals that Theodore Cooper and other blamed engineers were simply victims of confusing and ill formed contractual obligations. As with most modern disasters, including those in CS and AI, there is an underlying tension between individual agency and bureaucratic control with actual responsibility lying somewhere in between [47, 53].

We argue that quick ethics fixes, like ethics modules largely developed for and within computer science, are not a sufficient intervention to actually teach CS students of how ethical challenges get resolved in real world contexts. In actuality, an ethical challenge would not get resolved by an individual or even a group of technologists but will require an inherently interdisciplinary effort. Furthermore, this solutionist attitude of quick fixes in CS de facto displaces the knowledge of other qualitative oriented educators and researchers. The result is not only a feedback loop within computer science, but a continuing aggregation of disciplinary privilege that seeks to make computer scientists claim both "technical" and "social" expertises, the latter of which they do not actually have in depth.

2.3 The Myth of the Ethical Unicorn

The rise in training computer scientists to be ethics "experts" accompanies a long degradation of enrollment [70], opportunity [22], and funding [17] of social sciences and humanities. In research, focuses on fairness and justice, qualitative methods, and broader impacts are often regarded—at least, in the fleeting discourse of the social media public square, like Twitter and Reddit—as not "real" machine learning. The underrepresentation of liberal arts backgrounds and fairness research in the technology sector intersects with diversity disparities in computing [83], generally, and AI specifically [78]—given that women, BIPOC, and queer scholars come from more diverse (non-computer science) backgrounds and contribute more research on identity-based fairness issues in technology [41].

Given the privileged position of computer science, there is at times a tendency for the discipline to ascribe to itself a certain self-importance, falling into an assumed role as the expected saviour in resolving any presented crisis. If the engineer as saviour is one of the last contemporary manifestations of symbolic self-worth and self understanding in computer science, we could identify at least two general attitudes which shaped the engineering professions in the last century.

The first is the attitude of the paternal power of the engineer, which is defined by a claim to exert total technical competency doubled by a position of exhaustive moral authority grounded in social status or in a specially acquired wisdom [1].

The second is the figure of the engineer expert who claims total mastery over technical knowledge of their field of competency and, contrary to the paternal attitude, does not claim moral superiority or any special access to wisdom, but may dismiss any responsibility requiring engagement with topics outside their defined realm of expertise [2].

The new figure of the socio-technical expert, able to exhaustively solve the most intractable societal problems, is a recent development. In fact, this new rhetorical figure is another manifestation of the pioneer technologist able to break boundaries and successfully overcome any imaginable challenge while heroically gracing humanity with the fruit of their works. This endeavor results in the unreasonable expectation of creating "Ethics Unicorns", "full stack" developers [80] with a sprinkle of social awareness on top.

Of course, this isn't as simplistic as a displacement of liberal arts by computer science; there is also a level of responsibility liberal arts has in its disciplinary philosophies. As danah boyd, a principal researcher at Microsoft, wrote of social science: "Academic disciplines are brutally myopic, judgmental of anyone who chooses to explore a path of inquiry outside of the acceptable boundaries of the field"[11]. Extending training beyond traditional social science modes of inquiry, into computing, has also been largely missing from liberal arts disciplines. Instead, students must meander into computer science classes or rarer interdisciplinary ones, like newer offerings of computational social science [48] and digital humanities[45]. However, we have chosen to focus specifically on computer science, and AI, due to its emergent and out-sized power: computer scientists and machine learning engineers are having real world impact, socially, politically, and economically. In the context of addressing the challenges of managing this tangible and widespread influence, there will need to be an expansion and fundamental re-definition of not only what it means to do computer science but also a mandate to include varying disciplinary interactions as collaborators in getting to meaningful interventions. At present, it is the computer scientists that have the most leverage to take the action necessary to make this change.

[1]A paternalistic attitude in engineering "assumes that it is morally permissible for one rational person (without the other's informed consent) to decide significant aspects of the other's life because she believes herself at least as able to judge such things as the other is." [19, p. 104]

This type of paternalistic attitude of an engineer has one of its first historical formulations in John Locke's account of paternal power : "The power, then, that parents have over their children arises from that duty which is incumbent on them, to take care of their off-spring during the imperfect state of childhood. To inform the mind, and govern the actions of their yet ignorant nonage, till reason shall take its place, and ease them of that trouble, is what the children want, and the parents are bound to [...]" [42, p. 32]

[2]"Engineers shall perform services only in the areas of their competence.[...] Engineers, when serving as expert or technical witnesses before any court, commission, or other tribunal, shall express an engineering opinion only when it is founded upon adequate knowledge of the facts in issue, upon a background of technical competence in the subject matter, and upon honest conviction of the accuracy and propriety of their testimony." from American Accreditation Board for Engineering and Technology (ABET) Code Of Ethics Of Engineers: The Fundamental Principles[32, p.2]

3 MECHANISMS OF EXCLUSION: FINDINGS & EVIDENCE

The way in which exclusion is operationalized within communities is the result of a series of specific mechanisms of exclusion—that is to say, methods through which AI developers, and computer scientists more broadly, remove themselves from related disciplines as they theorize and operationalize approaches to develop new models.

Computer science (CS) and humanistic social science (HSS) exclude each other through specific accepted attitudes and norms, as well as practices. After qualitatively analyzing a survey of 254 AI ethics courses from 132 universities [3] — 151 explicitly marked as undergraduate courses and 92 marked as explicitly graduate courses — we were able to identify the following evidence for mechanisms of exclusion in AI ethics education. We note that of the analyzed courses, we were only able to review the full syllabus for 84 courses — as 167 courses did not have syllabi freely available online and 3 of the examined syllabi were not in English so not possible for the authors to review. The 84 syllabi, from 54 academic institutions, represented the lesson plans and readings for 58 undergraduate courses and 26 graduate courses.

The most well represented department in our survey sample by far is Computer Science, representing over 30% of analyzed courses (see Table 1). We also include courses taught in related STEM fields such as Mathematics, and Engineering. Altogether, these STEM departments host about 44% of the surveyed AI ethics courses. The HSS departments best represented in this survey are Information or Library Sciences, Philosophy, and Communications. We also include related HSS departments like Media Studies, Science Technology and Society (STS), and professional departments like Law and Business. Courses in these dominant HSS-skewed departments make up about 52% of the surveyed dataset. An additional 3% of the courses are located in miscellaneous "Other" departments (ie. History, Governance and Policy, Journalism, Medicine or other sciences, English, Film, Urban studies, Design, Sociology). Only 1% of the surveyed AI ethics courses were developed for explicitly interdisciplinary departments.

From our analysis, we identified that exclusion manifests in mainly in the following ways: methodological dogmatism, a lack of collaborative and interdisciplinary outputs and siloed citations.

3.1 Methodological Dogmatism

Methodological dogmatism is to say that the disciplines *don't value each other*, and more specifically, do not value their respective processes for arriving at an accepted conclusion or "truth". Different disciplinary approaches to the topic of ethical considerations in AI

[3]Details of our analysis can be found here. We thank Casey Fiesler for her efforts in crowd-sourcing the resource we used as the basis of our analysis.

Table 1: Overall AI Ethics Courses Analyzed by Department

Discipline	Number of Courses
83	Computer Science
54	Information or Library Sciences
31	Philosophy
21	Engineering
21	Communications
9	Law
7	Mathematics
7	Media Studies
6	Other
6	Business
5	STS
4	Interdisciplinary

are often in contempt of each other and unappreciative of the norms and methodologies used by another group. As a result, a dismissal on both sides ensues. It is rare to see any shared methodology or a “mixed method” approach involving more than one disciplinary lens.

For instance, quantitative methods can be looked down upon by those coming from a more holistic approach—it is taught in some HSS courses that quantitative interventions are insufficient to address certain harms, mainly because to arrive at such an intervention requires an uncomfortable amount of abstraction, which removes the true complexity of the situation. On the other hand, several AI ethics courses taught in CS departments overwhelmingly feature analyses of ethical issues from a purely computational perspective or with the theoretical and engineering methods familiar to the STEM disciplines. A few of the analyzed CS courses actually had a primary emphasis on the completion of a set of defined technical projects or problem sets, rather than including readings or scheduled class discussions.

Several courses do not even indicate explicitly the need for cross-disciplinary collaboration, and most do not provide tools or strategies to educate students on how to successfully seek and navigate such collaborations. Only one of the surveyed courses explicitly mentions disciplinary variance in methodology in their syllabus. This course is also the only one analyzed that presented students with methodological approaches from HSS to supplement computational thinking. None of the surveyed courses explicitly mention as part of the curriculum a discussion of the limitations of their own disciplinary approach in addressing the issue of AI ethics.

3.2 Lack of Joint Outputs

The other mechanism of exclusion evident in this space is that of a lack of effort to engage in interdisciplinary translation. This is to say that there is an apparent acceptance that the disciplines *don’t talk to each other*, and effectively tend to not communicate across disciplinary lines. Common terms in the AI ethics community can be interpreted and represented completely differently from various disciplinary lenses at the same conference, with confusion manifesting in the inherent incompatibility of certain employed definitions and perspectives [82].

In terms of pedagogy, there does not seem to be an attempt to shift towards a shared vocabulary. Of the 254 surveyed courses, there were only 5 instances of courses allowing for cross-disciplinary teaching or open courses with non-prohibitive pre-requisites, which allowed students from various majors to take the course together. Only one course was explicit in the need for no pre-requisites as a requirement to keep the material accessible to any interested student.

From the syllabi it is also clear that the courses are made for a homogeneous audience. In the majority of cases, suggested readings are full of disciplinary references and technical terms anchored to assumed knowledge. In HSS derived courses, vocabulary is fairly inaccessible by virtue of being overly specific or referential to specific theories. For example, the ethical theories of philosophy, often taken as assumed knowledge in HSS courses, may be unfamiliar for many STEM students. In the same way, the CS courses do not often invest in translating equations or programming details to make the coursework accessible and bridge communication hurdles. In addition to this, professional ethics courses will often reference industry-specific ethics statements such as the ACM Code of Ethics [81] or familiar textbooks such as "Artificial intelligence: a modern approach" [66].

35 of the surveyed courses were a required course—mostly for computer science students but also for related majors such as data science, business computing or information science. The Engineering Accreditation Commission (ABET) requires an ethics course as part of their accreditation process for a program. As a result, since every CS student has to take such courses anyways, it may not make sense to open such a course for a broader audience. There are at times, several AI ethics courses run independently by departments at the same school—for instance, University of California Berkeley has AI Ethics courses run out of it’s Law School, Information School, Computer Science and History departments. Only 4 surveyed courses were hosted by an explicitly interdisciplinary effort—often run by an institute at the university working with multiple departments.

This lack of effective translation poses a clear threat on the ability of students to get the opportunity to co-author or co-create with others of a different disciplinary lens. Although researchers from different disciplinary origins may present to each other or consult on each other’s papers, if attention is not paid to establishing a common understanding then these types of exchanges are ultimately performative rather than substantial. The frustration of translation often prevents a shared ground to properly engage at the level of adequate mutual contribution as co-authors and primary collaborators. It is clear that the pedagogical norms in AI ethics courses often does little to prepare students to overcome that challenge.

3.3 Siloed Citations

Finally, as a side effect of both of these previously identified mechanisms of exclusion, we can see a lack of cross-citation—ie. *a lack of effectively building on each other’s work*. The lack of cross-citation across disciplinary lines indicates that findings from one discipline are not respected as adequate evidence to inform researchers from another and that, even if valued, the insufficient accessibility of knowledge across disciplines may make results too difficult to be properly understood and engage with.

Textbooks and articles written by authors of a CS background are rarely included in the syllabi of AI ethics courses run by HSS departments, including departments of Information Sciences, Communication and Media Studies. Among the most often cited work from HSS syllabi is "Digital Community, Digital Citizen" by Jason Ohler [57], "Algorithms of Oppression" by Safiya Umoja Noble [56], "Automating Inequality" by Virginia Eubanks [27] and "Digital Media Ethics" by Charles Ess [26]. All of these authors work in HSS departments or have an HSS educational background. Similarly, frequently referenced work in CS curricula includes "Ethics for the Information Age" by Michael J. Quinn [63], "Superintelligence: Paths, Dangers, Strategies" by Nick Bostrom [10], and "A Gift of Fire" by Sara Baase & Timothy M. Henry [4]—all of whom have either a CS background or academic CS affiliations. Popular references also include books not written by computer scientists but still favoring a more analytical or economic interpretation of issues, such as "Weapons of Math Destruction" [58], or "The Second Machine Age"[12]. From the survey, there were few if any examples of featured textbooks written by co-authors of differing disciplinary backgrounds.

This disciplinary divide seems to correlate to fundamentally differing views of the challenges inherent to building algorithms ethically. HSS anchored literature tends to focus much more heavily on the human interactions of those making decisions regarding algorithmic systems and those impacted, while CS anchored literature generally frames issues as challenges in engineering responsibility or system control in the presence of unintended consequences.

It is noteworthy that, as a relatively new topic to present to students, much of the assigned material across all courses included a curated set of online media highlighting studied cases. Popular examples include investigative journalism piece "Machine Bias" by Propublica, and a short film titled "Humans Need Not Apply" directed, produced, written and edited by CGP Grey.

4 FORMATION OF EXCLUSIONARY PEDAGOGY & CONSEQUENCES OF EXCLUSION

As shown above, this crisis of recognition can be understood from practices of disciplinary devaluation, as a refusal to create common vocabularies and from a rejection to form enduring joint bodies of knowledge. One question that we need to ask at this point is: what are the reasons for this systematic indifference, devaluation and lack of recognition between CS and HSS and what are the consequences? What reasons prevent a sustained and systematic collaboration between HSS and CS, and how does this lack of interaction hinder our ability to address interdisciplinary challenges such as those in AI ethics?

4.1 The Inherent Exclusion of Disciplinary Classifications

Commonly accepted descriptions of disciplinary classifications are not merely neutral and objective. In fact, they enforce and reproduce the mechanisms of exclusion that we outlined above.

Disciplines are defined in two ways [62]: by their own epistemologies, that is their own bodies of knowledge—including concepts, methods and objective aims [73]—or as the "organised social groupings" through which they operate [79]. Embedded in this definition is the explicit axiom that one cannot separate the production of knowledge understood as content from the social actors and their conduct.

TABLE I. Broad disciplinary groupings

Biglan	Kolb	Disciplinary areas
Hard pure	Abstract reflective	Natural sciences
Soft pure	Concrete reflective	Humanities and social sciences
Hard applied	Abstract active	Science-based professions
Soft applied	Concrete active	Social professions

Figure 1: Table of a taxonomy for disciplinary groupings, according to Kolb. Source: Becher, 1994 [6].

Kolb [46] and Biglan [8], articulate four main axes to frame disciplinary groupings (see Figure 1 and 2, respectively). While Biglan's focus is on identifying the subject matter of disciplines, Kolb draws attention to the mode of inquiry operative in the classified disciplines. Thus, Kolb describes disciplines in terms of "abstract reflective", "concrete reflective", "abstract active" and "concrete active" [46]. Meanwhile, Biglan's disciplinary classification, "based upon the perceptions of academics", involves an investigation on how a "panel of 168 university faculty members viewed a sample of 36 disciplines" [8]. According to this study, academics view disciplines along two key axes—"hard" vs. "soft" and "pure" vs. "applied"—which in turn engenders variations of their own: "hard pure", "soft pure", "hard applied" and "soft applied" [8].

We observe that disciplinary classifications, analyzed at an axiological level, enforce and reproduce disciplinary biases, most importantly providing justification and legitimacy for the formation of mechanisms of exclusionary pedagogy in CS. If we look at how CS is labeled according to Biglan's classification, we can see that it operates as a "hard" and "applied" discipline. Kolb for their part sees CS as "abstract" and "active". In stark contrast, HSS have almost a virginal quality being labeled in the exact opposite categories, as being irrevocably "soft" and "pure" (Biglan), "concrete" and "reflective" (Kolb).

We hypothesize that behind these constructed distinctions of "hard" vs. "soft" or "active" vs. "reflective", there are serious assumptions about the nature of academic problems, historically anchored to economic beliefs which shape who is worthy and who is not. It is the distinction between accuracy and inaccuracy, between tangibility and speculation, between clear and distinct on one end and confused and ambiguous on the other. The former state is understood to be much more real, and valuable, being framed as more economically relevant.

If our hypothesis is correct, this disciplinary classification is problematic in the context of CS pedagogy, which may be categorized as "hard" or "active" to indicate its economic relevance, but actually operates in conditions of epistemic vagueness. Despite being a nominally "hard" discipline, CS requires a *plurality* of perspectives and modes of analysis, involving both the consideration of abstract theory and engineering, in addition to the challenge of handling both the practical and conceptual ambiguity typically characteristic of nominally "soft" disciplines.

It’s thus clear that disciplinary classifications draw their legitimacy from internal modes of recognition operative in the disciplines themselves and from the nature of the impact they have on societies which validate or contest their self-understanding. These disciplinary classifications should be understood as essentially contingent and could be formulated differently than they actually are. The positioning of HSS in stark contrast and opposition to CS is thus an imagined dichotomy. In reality, each discipline shares more in common than previously imagined, with several intersecting challenges. Setting up false dichotomies reinforces a false belief in the existence of inherently opposing, rather than complementary, disciplinary norms and ultimately escalates in the observed reluctance of differing disciplines to mutually engage.

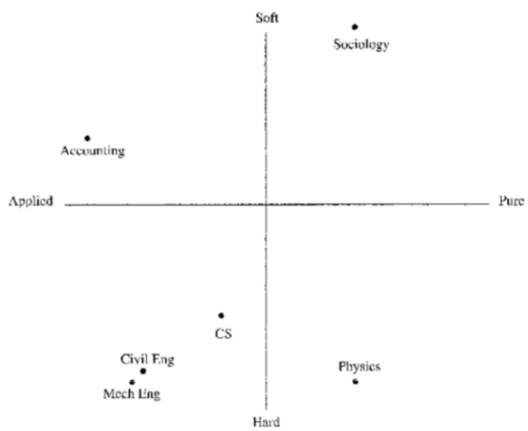

Figure 2: Selected disciplines located on the main Biglan dimensions. Source: Clark, 2003 [15].

4.2 CS Pedagogy and Hierarchies of Knowledge

Disciplinary classifications display profound and entrenched disciplinary biases according to which some modes of knowing are presented as sure, robust and legitimate at the expense of other modes of knowing seen as vague, unwarranted, and superfluous. This establishes the perceived hierarchy of knowledge that leads to exclusionary pedagogy and practice.

In this framing of disciplines based solely on subject matter (ie. content and method), the natural sciences, engineering and computer science operate as "hard" or "paradigmatic" whereas the humanities are seen as "idiosyncratic". These latter disciplines do not qualify as "paradigmatic" and are classified as *irreversibly* "soft". The social sciences, such as sociology and business fields are thus framed as being in futile search of a paradigm. That is, they display a potentiality for finding a paradigm but have not been able to find one and therefore are categorized as *relatively* "soft" to the STEM disciplines.

The second dimension of Biglan’s classification is defined by the relation between "applied" and "pure" aspects of disciplines. Accounting and engineering fields, including CS, are thus implicitly labeled as practical, whereas the natural sciences, social sciences and the humanities are categorized as abstract, implying less tangible utility [4]. The label of "active" being applied to STEM disciplines also reinforces the false assumption that CS, as an "active" discipline, is meant to hold the agency in situations requiring intervention and possesses the greater influence on system and societal outcomes. The framing of CS as being opposite to the passive category of "reflective", also devalues the importance of reflection in responsible CS development, while falsely framing reflection in opposition to effectiveness and efficiency.

Despite these small variations in terminology, we contend that the disciplinary classifications are not simply descriptive. In our view, these disciplinary labels are essentially evaluative and have social effects in terms of symbolic and economic worth, social relevance and impacts. In fact, they work as markers of truth which carry and reproduce historical and social hierarchies of knowledge. In a word, they justify the dominance and exclusion of “soft” modes of knowing by grounding underlying disciplinary values which in turn allow for a restricted set of choices, commitments and actions within given disciplines.

Interest	Knowledge	Medium	Science
Technical	Instrumental (causal explanation)	Work	Empirical-analytic or natural sciences
Practical	Practical (understandin g)	Language	Hermeneutic or 'interpretive' sciences
Emancipatory	Emancipatory (reflection)	Power	Critical sciences

Figure 3: Taxonomy of Disciplines from Habermas. Source: Clear, 2001 [16].

As a way for us to see the limits and bias operative in disciplinary classifications around "hard", "soft", "pure", "applied", "active" or "abstract" framings, we should compare them with a different classification schema of disciplines. Take, for example, Habermas’ disciplinary classification [40]. If we follow Habermas (see Figure 3), CS would be exclusively defined by technical interests, instrumental knowledge and causal explanations, and work within the orbit of empirical or natural sciences. HSS, however, will belong to the sphere of interpretive, critical sciences and be described by it’s methodological tendency to mobilize practical and emancipatory knowledge via the medium of language and power.

Neither of these categories appears contradictory or possess connotations of inherent relative value. Habermas’ disciplinary

[4]Despite Biglan’s label of computer science as applied there is debate in CS of whether such a label is warranted. For an alternative view of computer science as a theoretical science that is both a pure and hard discipline see Dijkstra’s 1989 comments: “And what is a science of computing about? Well, when all is said and done, the only thing computers can do for us is to manipulate symbols and produce results of such manipulations. From our previous observations we should recall that this is a discrete world and, moreover, that both the number of symbols involved and the amount of manipulation performed are many orders of magnitude larger than we can envisage: they totally baffle our imagination and we must therefore not try to imagine them.” [21]

classification establishes as valuable the proposition which states that knowledge is not disinterested but is dependent on interests and by consequence is perspectival and partial. If we ask ourselves what are the assumptions inherent in interests operative in Biglan's disciplinary classification, we would agree that technical interest based in causal explanations define the "hard" disciplines. However, we would be hard pressed to find any mention to grant scientific status, robustness or accuracy to interpretive analysis and emancipatory goals defining what they call the "soft" disciplines. Under this classification scheme, it becomes clear that both disciplines simply differ and overlap in certain complementary ways that assert the nature of each perspective's contribution, rather than framing one discipline as superior to another.

Hierarchies of knowledge contained in disciplinary classifications thus carry, reproduce and enforce specific epistemological assumptions and commitments. As we can see, disciplinary classifications which define CS as "hard" and "applied" or as "abstract" and "active" do so at the expense of other modes of knowledge labeled as soft, idiosyncratic and un-paradigmatic. If, however, we choose to frame disciplines differently, in the classroom and within our own work, then it is evident that much of the rhetoric feeding into the current state of disciplinary exclusion is not fixed or absolute, and can actually be revisited and rejected, leaving open the new possibility of more inclusive disciplinary classifications and practice.

4.3 Consequences of Exclusion

If we agree that disciplinary classifications reproduce hierarchies of knowledge which are in essence exclusionary, what are the consequences of the mechanisms of exclusion that we analyzed above?

One obvious response is *disciplinary self-isolation*, and the resultant limit to growth based on the existing restrictive values and rigid assumptions that animate the discipline thus far.

However, we can further articulate our answer in the form of a paradox that is at the heart of CS: the discipline finds itself in this space where its technical artifacts have a global reach, yet its methods of analysis, concepts, values and assumptions reflect only a narrow subset of the social world that it finds itself embedded in. Thus, in a way, CS as a discipline is inflicted with a crisis of representation that is revealed by the narrow scope of its techniques and the modes of analysis that shape the discipline.

The continuation of this narrow disciplinary approach, which reproduces the aforementioned mechanisms of disciplinary exclusion, is not only costly but also harmful. If CS as a discipline continues this trend, we need to reflect on the consequences of not engaging with other disciplines and in particular the social costs of not engaging with HSS. To put it differently—what are the values, interests and goals that guide HSS and are lacking in CS? If we follow Habermas, HSS is framed around humanistic goals of emancipation, participaction, and respectful inclusion of all stakeholders present or absent, powerful or not. These are thus the priorities that the CS field will fundamentally neglect once they disengage with HSS. Such a disassociation would be tragic, given that these are all considerations that represent key ideals, commitments and values for responsible CS ethics development. The consequences of this exclusion thus involve not only disciplinary self-isolation, but, most importantly, entails *a loss of values, assumptions and methods* that are crucial in HSS: hermeneutical, interpretive, qualitative methodologies and a sustained reflection on emancipatory societal goals.

The study of climate change could constitute a normative analogue from which CS pedagogy could find inspiration. Recent research shows that tackling climate cannot stand on a narrowly constructed group of experts. Rather, climate change education requires a holistic [5] theoretical [6] framing and a plurality of epistemologies, methodologies, instruments of analysis perspectives, values and assumptions. In a word, since climate change is a global and complex problem, it requires a pluralistic formulation of solutions [7]. Several examples of such a disciplinary shift can now be found—for instance, in a study on significant climate change, the Arctic researchers showed how "Sami reindeer herder interviews and observational weather data" formed qualitative studies which measured community perception of climate change and complemented "quantitative assessments of trends in temperature and precipitation." The objective "was to use a methodological approach encompassing the complexity, subjectivity, and context-dependent high sensitivity usually associated with qualitative methods; along with the scale, consistency, generalizability, and validity more generally associated with quantitative methods [...]"[34].

Given that CS technical artifacts have a global impact on our lives, accuracy understood as a techno-scientific requirement of clarity and distinctiveness is not enough. Applied blindly, irrespective of context and social power imbalances, accuracy as a practice, value and assumption is, in fact, a formula for stabilizing forms of dominance as it amplifies and reproduces historical and structural inequities. Instead, we need to focus on a critical engagement with exclusionary disciplinary classifications and their consequences in order to develop a new ethics of collaborative pedagogy at the intersection of CS and HSS.

5 TOWARDS A COLLABORATIVE PEDAGOGY

Given these mechanisms of exclusion operative in CS and AI inflected pedagogy, the final question to address is this: what should we do now? So far, we have seen that disciplinary norms create a crisis of interdisciplinary recognition which has heavy costs. Crisis is understood here as a moment of decision in light of the fact that CS technical artifacts permeate and impact almost every aspect of our lives. This crisis requires an expansion to the default framing

[5] In a paper titled "Towards more predictive and interdisciplinary climate change ecosystem experiments" which discusses "how computational and technological advances can help in designing experiments" the authors conclude: "We foresee that the holistic approach outlined in this Perspective could yield more reliable, quantitative predictions of terrestrial ecosystem response to climate change, and could improve knowledge on the value of ecosystem services and their links with ecosystem processes." [64]

[6] "A diversity of theoretical perspectives and methodologies is valuable to shine light on [climate change] from different angles; however, it would be beneficial to the field to do some of the intellectual heavy-lifting that might result in an integrative theory, that spans across the theoretical perspectives. In keeping with our introduction, we come together in conversation and reflection to pose our hopes for the field moving forward."[13]

[7] Such solutions can be conceived as research paradigms [18] or a question of the epistemology informing interventions: the "form and nature of knowledge and what can be known about it" [39] or "ways of knowing" [59].

of both CS and HSS, though there is ideological resistance in both disciplines for such a shift to practically occur.

We contend that adequately tacking these issues will require changes to the constitution of how we educate students hoping to one day address the crisis. Specifically, we hope that demonstrating the necessary shift from current exclusionary norms towards a more collaborative pedagogy in the classroom will facilitate the development of more inclusive instruction and participation in the field, for both academic and industry contexts.

5.1 Transversality in AI Ethics

CS pedagogy on its own is not able to elaborate the disciplinary norms and create conditions for stable comprehensive and socially beneficent technical artifacts. In other words, the creation of theories, tools and methods in AI ethics need to be understood as *transversal problems*, involving methods, theories and collaborators across several traditional disciplines. Not only would this new CS pedagogy play a crucial role in addressing the divide between CS and HSS but it would focus on developing solutions in a different way from our current default of developing methods and technical artifacts first and only weighing their impacts later—it would allow for technologists to imagine their role as a collaborative one with peers whose perspective and approach could be key to addressing what continues to be major challenges in the field.

The power of acting in concert in CS pedagogy would involve the acknowledgement of a requirement of *transversality*. This requirement is a blind spot both in the design of the collaborations themselves and in the way we diagnose problems. The field of CS has not yet come to the full realization that it deals with problems which exceed its traditional field of competency and in fact its problems are not merely technical or moral problems, but they are in fact transversal problems which require a diverse set of skills, and methodologies. A transversal problem is distinct from an interdisciplinary problem as its solution is not found in-between given disciplines but should be constructed from the effects on the stakeholders that could be or were impacted by it, and from a critique of the types of formal and substantive assumptions, choices, requirements and methodologies that are currently built into AI ethics pedagogy.

5.2 A Path Forward

The starting questions for us to develop actionable items in CS pedagogy are as follows: how can we build a curriculum for both social science and computer science to tackle technical issues, together? How can we develop a AI ethics curriculum that embraces transversality?

First, we recommend to focus on thinking and acting differently by including broad non-CS expertise and researchers when dealing with technical artifacts which have clear social impact. Demonstrate examples of effective collaborative outcomes to students—in the form of papers, books or other media effectively co-authored by scholars of different disciplinary perspectives, advocacy campaigns done in conjunction with affected communities, or projects making meaningful use of mixed methods. If possible, we suggest to encourage or allow students to tangibly collaborate across disciplines as part of their required outcome for the class. This can be made easier by running concurrent or joint courses with other departments, and minimizing prohibitive prerequisites, so that students may gain practical project experience engaging with collaborators of a different disciplinary lens.

Second, we need to educate students on frameworks of intervention based around existing problems, not anchored to the existing skills of those assumed to be in the position to address the problems. Real world ethics problems call for a diverse set of skills. Educators should focus not only on developing the technical skills or social theory skills of students. Instead, more attention should be paid to the value of appropriately articulating the right problem, as well as acknowledging and engaging the right stakeholders. This can be learnt through a frequent analysis of concrete case studies, complete with clear examples of the required contribution of those participating from various perspectives to address the studied situation. This can also be addressed by creating moments of joint collaborative spaces between students of varying disciplinary backgrounds or including as part of the educational process instruction on the methodology of other disciplines. This serves not to necessarily *expand* the toolbox of student skills, but rather *expose* them to the methodologies and vocabularies they will need to recognize, awknowledge and respect as part of the collaboration process with peers of other disciplines.

Third, we need to incorporate explicit references in the AI ethics syllabi of stakeholders beyond the technologist, including discussions of their roles and how they are impacted. One of the key issues that a syllabus should address is a more inclusive identification of stakeholders when formulating an AI problem. One of the questions that every AI syllabus should account for is not only who the target audience is but who the impacted parties may be. Beyond institutional stakeholders, such as product managers or engineers, there could be greater discussion of societal stakeholders, such as policymakers and regulators, as well as more speculative reflection on who perhaps *could* contribute to the solution, though may be rarely invited to do so. Different disciplines have different approaches to including and incorporating the perspective of stakeholders and mapping this relationship out but it is critical to present at this point the need for collaboration to address multi-faceted, complex problems rather than embrace the myth of a sufficiency in just expanding technical expertise to include some ethical understanding.

Fourth, we need to develop frameworks to work with affected populations and experiential experts, including community organizer toolkits and speakers. We could imagine an AI ethics class, for example, in which the narration or participation of experiential experts is heavily featured, and students can share their own reflections on lived experiences with unjust algorithms, thus connecting personal testimonies to broader studied accounts of the social and ethical implications of deployed or speculative systems. The account of practitioners could be included in the class to explain the logics informing how technical artifacts are designed, and the kind of affordances or limitations present. Such exercises build empathy towards the variety of perspectives present on these issues, and encourage an open-mindedness to learn from and seek out an alternative lens.

Fifth, we suggest engaging students in the exercise of assessing disciplinary competence in a range of situations and developing the ability to identify the relevance of a particular approach to address challenges for specific types of problems or in particular situations. One may ask questions such as: When is theoretical analysis needed? What are the dilemmas and aporias embedded in the problem at hand? Why are methods and patterns of quantification needed in this context? There should be an explicit conversation about not only what is gained from the methods embraced by one's discipline, but also what is lost—and how the limitations of one's disciplinary lens can be addressed by looking towards mixed method approaches, or knowledge and collaboration with other disciplines to fill in certain ability gaps.

6 CONCLUSION

Working towards an effective ethics of responsibility from a technological perspective involves a steady procession towards a future of one primary goal—to maintain the permanent conditions of our collective existence and well being. To put it in starker terms, our pedagogical design efforts should not divorce our epistemological concerns for truth, accuracy, robustness, and metrics from our ethical concerns to achieving the public good, collective responsibility, shared practical reasoning and social justice. As a way to address this structural divide operative in the ways ethics is taught in computer science, we need to develop pedagogies able to forge a new ground for the relation between epistemology and ethics, truth and the good, individual and collective responsibility.

Computer science is not the only community to face this challenge. Similar acknowledgment in climate science of the inability of purely quantitative thinking led to its innovation to embrace other forms of knowledge and a wider array of methodology in its teaching practice. By moving beyond just the measured artifacts of the climate change crisis, and embracing tools from economics, anthropology, behavioural psychology, narrative storytelling and more, the community was ultimately more effective in advocating for its cause and garnering impactful outcomes beyond what the original meterologists and ecologists were in a position to address. With effort and more conscious planning, we can only hope to see the same for our own crisis of ethics in AI.

7 CONTRIBUTION STATEMENT

This was a largely collaborative work. Each author regularly met to discuss the paper and formulate the argument. Still, we wish to lay out what each author did in this work for transparency. The first author did the majority of the writing and editorial work, as well as the empirical analysis. She also conducted the final editorial work on the paper, including writing, synthesizing, and formatting. The second author worked to shape the larger conceptual frame of the paper. He also conducted a large portion of the literature review foundation to the argument. The third author input social and political theory perspectives, aided in imagining future directions of the work, and contributed to editorializing the work.

8 POSITIONALITY STATEMENT

Reflexivity in research has become established practice in anthropology ([65]), sociology ([3]), and feminist studies ([25])—and increasingly, computing domains like human-computer interaction (e.g., [69]). Each of the authors on this work have been challenged to engage with their own experiences and backgrounds in researching the interdisciplinary space of FAccT. The authors perspectives are shaped by a collection of backgrounds in: social and political theory, engineering, computer science, gender studies, and human-computer interaction. This collection of disciplinary backgrounds has shaped both our pragmatic and philosophical approaches to computing approaches. Our experiences as students and teachers, at different aspects of our careers, has led us to care deeply about educational approaches in computing disciplines—including how to embrace new approaches for preparing future students on handling ethical dilemmas. While we each have histories in different countries, each author's education is rooted in Western scholarship, which has shaped our approach to AI ethics pedagogy. We each have our own set of privileges and marginalizations that situate not only our ways of conducting research, but our intentions and motivations in doing so, as well as shaped our understanding of the damage inflicted by any exclusionary practice. Each of us have also worked with large technology companies headquartered in the United States as well as spent time in American academic institutions.

REFERENCES

[1] Philip E Agre. 1997. Lessons learned in trying to reform AI. *Social science, technical systems, and cooperative work: Beyond the Great Divide* 131 (1997).
[2] Ronald E Anderson. 1992. ACM code of ethics and professional conduct. *Commun. ACM* 35, 5 (1992), 94–99.
[3] Margaret Archer. 2007. *Making our way through the world: Human reflexivity and social mobility.* 1–343 pages. https://doi.org/10.1017/CBO9780511618932
[4] Sara Baase. 2012. *A gift of fire.* Pearson Education Limited.
[5] Theresa Beaubouef and John Mason. 2005. Why the high attrition rate for computer science students: some thoughts and observations. *ACM SIGCSE Bulletin* 37, 2 (2005), 103–106.
[6] Tony Becher. 1994. The significance of disciplinary differences. *Studies in Higher education* 19, 2 (1994), 151–161.
[7] Angela R Bielefeldt, Madeline Polmear, Daniel Knight, Christopher Swan, and Nathan Canney. 2017. Ethics Across the Curriculum? Integrating Ethics and Societal Impact Topics into Core Engineering Courses. In *Proceedings of the ASEE American Society for Engineering Education Rocky Mountain Section Conference.* 1–15.
[8] Anthony Biglan. 1973. The characteristics of subject matter in different academic areas. *Journal of applied Psychology* 57, 3 (1973), 195.
[9] Judith A Boss. 1994. The effect of community service work on the moral development of college ethics students. *Journal of Moral Education* 23, 2 (1994), 183–198.
[10] Nick Bostrom. 2017. *Superintelligence.* Dunod.
[11] Danah Boyd. 2016. Why social science risks irrelevance. *The Chronicle Review* (2016).
[12] Erik Brynjolfsson and Andrew McAfee. 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* WW Norton & Company.
[13] KC Busch, Joseph A Henderson, and Kathryn T Stevenson. 2019. Broadening epistemologies and methodologies in climate change education research. *Environmental Education Research* 25, 6 (2019), 955–971.
[14] Martyn Clark. 2003. Computer Science: a hard-applied discipline? *Teaching in Higher Education* 8, 1 (2003), 71–87.
[15] Martyn Clark. 2003. Computer Science: a hard-applied discipline? *Teaching in Higher Education* 8, 1 (2003), 71–87.
[16] Tony Clear. 2001. Research paradigms and the nature of meaning and truth. *ACM SIGCSE Bulletin* 33, 2 (2001), 9–10.
[17] Code.org. 2019. US Dept of Education Prioritizes Computer Science Funding.
[18] John W Creswell. 2003. A framework for design. *Research design: Qualitative, quantitative, and mixed methods approaches* (2003), 9–11.
[19] Michael Davis. 1998. *Thinking like an engineer:Studies in the Ethics of a Profession.* Oxford University Press New York, NY, USA. 143 pages.
[20] Peter J Denning. 1989. A debate on teaching computing science. *Commun. ACM* 32, 12 (1989), 1397–1414.
[21] Edsger W Dijkstra et al. 1989. On the cruelty of really teaching computing science. *Commun. ACM* 32, 12 (1989), 1398–1404.

[22] Erin Duffin. 2020. Projected starting salaries for Bachelor's degree graduates in the United States in 2020, by discipline.
[23] Amnon H Eden. 2007. Three paradigms of computer science. *Minds and machines* 17, 2 (2007), 135–167.
[24] Amnon H Eden. 2007. Three paradigms of computer science. *Minds and machines* 17, 2 (2007), 135–167.
[25] Kim England. 1994. Getting Personal: Reflexivity, Positionality, and Feminist Research. *The Professional Geographer* 46 (01 1994), 80 – 89. https://doi.org/10.1111/j.0033-0124.1994.00080.x
[26] Charles Ess. 2013. *Digital media ethics*. Polity.
[27] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
[28] Casey Fiesler, Natalie Garrett, and Nathan Beard. 2020. What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 289–295.
[29] Casey Fiesler, Natalie Garrett, and Nathan Beard. 2020. What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 289–295.
[30] Allan Fisher and Jane Margolis. 2002. Unlocking the clubhouse: the Carnegie Mellon experience. *ACM SIGCSE Bulletin* 34, 2 (2002), 79–83.
[31] James Fleck. 1994. Knowing engineers?: A response to Forsythe. *Social studies of science* 24, 1 (1994), 105–113.
[32] Accreditation Board for Engineering and Technology(ABET). 1977. Code of Ethics of Engineers. (1977).
[33] Diana Forsythe. 2001. *Studying those who study us: An anthropologist in the world of artificial intelligence*. Stanford University Press.
[34] Maria Furberg, David M Hondula, Michael V Saha, and Maria Nilsson. 2018. In the light of change: a mixed methods investigation of climate perceptions and the instrumental record in northern Sweden. *Population and environment* 40, 1 (2018), 47–71.
[35] Howard Gardner. 1991. The tensions between education and development. *Journal of Moral Education* 20, 2 (1991), 113–125.
[36] Jamie Gorson and Eleanor O'Rourke. 2019. How Do Students Talk About Intelligence? An Investigation of Motivation, Self-efficacy, and Mindsets in Computer Science. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*. 21–29.
[37] Daniel Greene, Anna Lauren Hoffmann, and Luke Stark. 2019. Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
[38] Barbara J Grosz, David Gray Grant, Kate Vredenburgh, Jeff Behrends, Lily Hu, Alison Simmons, and Jim Waldo. 2019. Embedded EthiCS: integrating ethics across CS education. *Commun. ACM* 62, 8 (2019), 54–61.
[39] Egon G Guba, Yvonna S Lincoln, et al. 1994. Competing paradigms in qualitative research. *Handbook of qualitative research* 2, 163-194 (1994), 105.
[40] Jürgen Habermas. 1979. *Communication and the Evolution of Society*. Vol. 572. Boston: Beacon Press.
[41] Peter L Hinrichs. 2015. Racial and ethnic differences in college major choice. *Economic Trends* (2015).
[42] Locke John. 1980. *Second treatise of Government*. Charles River Editors.
[43] Deborah Johnson. 1994. Who should teach computer ethics and computers & society? *ACM SIGCAS Computers and Society* 24, 2 (1994), 6–13.
[44] Khari Johnson. 2020. A fight for the soul of machine learning. (2020).
[45] Steven E Jones. 2013. *The emergence of the digital humanities*. Taylor & Francis.
[46] David A Kolb. 1981. Learning styles and disciplinary differences. *The modern American college* 1, January1981 (1981), 232–235.
[47] Eda Kranakis. 2004. Fixing the blame: Organizational culture and the Quebec Bridge collapse. *Technology and culture* 45, 3 (2004), 487–518.
[48] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. 2009. Social science. Computational social science. *Science (New York, NY)* 323, 5915 (2009), 721–723.
[49] Colleen Lewis, Paul Bruno, Jonathan Raygoza, and Julia Wang. 2019. Alignment of goals and perceptions of computing predicts students' sense of belonging in computing. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*. 11–19.
[50] C Dianne Martin. 1997. The case for integrating ethical and social impact into the computer science curriculum. In *The supplemental proceedings of the conference on Integrating technology into computer science education: working group reports and supplemental proceedings*. 114–120.
[51] C Dianne Martin, Chuck Huff, Donald Gotterbarn, and Keith Miller. 1996. Implementing a tenth strand in the CS curriculum. *Commun. ACM* 39, 12 (1996), 75–84.
[52] C Dianne Martin and Elaine Yale Weltz. 1999. From awareness to action: Integrating ethics and social responsibility into the computer science curriculum. *ACM SIGCAS Computers and Society* 29, 2 (1999), 6–14.
[53] C Dianne Martin and Elaine Yale Weltz. 1999. From awareness to action: Integrating ethics and social responsibility into the computer science curriculum. *ACM SIGCAS Computers and Society* 29, 2 (1999), 6–14.
[54] Ralph Nader. 1965. *Unsafe at any speed: The designed-in dangers of the American automobile*. New York: Grossman.
[55] Norman R Nielsen. 1972. Social responsibility and computer education. *ACM SIGCSE Bulletin* 4, 1 (1972), 90–96.
[56] Safiya Umoja Noble. 2018. *Algorithms of oppression: How search engines reinforce racism*. nyu Press.
[57] Jason B Ohler. 2010. *Digital community, digital citizen*. Corwin Press.
[58] Cathy O'neil. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
[59] Michael Quinn Patton. 2002. Two decades of developments in qualitative inquiry: A personal, experiential perspective. *Qualitative social work* 1, 3 (2002), 261–283.
[60] Arnold Pears, Stephen Seidman, Lauri Malmi, Linda Mannila, Elizabeth Adams, Jens Bennedsen, Marie Devlin, and James Paterson. 2007. A survey of literature on the teaching of introductory programming. In *Working group reports on ITiCSE on Innovation and technology in computer science education*. 204–223.
[61] Cynthia Pearson and Norbert Delatte. 2006. Collapse of the Quebec bridge, 1907. *Journal of performance of constructed facilities* 20, 1 (2006), 84–91.
[62] Derek J Price. 1970. Citation measures of hard science, soft science, technology, and nonscience. *Communication among scientists and engineers* 3 (1970), 22.
[63] Michael J Quinn. 2014. *Ethics for the information age*. Pearson Boston, MA.
[64] Francois Rineau, Robert Malina, Natalie Beenaerts, Natascha Arnauts, Richard D Bardgett, Matty P Berg, Annelies Boerema, Liesbeth Bruckers, Jan Clerinx, Edouard L Davin, et al. 2019. Towards more predictive and interdisciplinary climate change ecosystem experiments. *Nature climate change* 9, 11 (2019), 809–816.
[65] Jennifer Rode. 2011. Reflexivity in digital anthropology. 123–132. https://doi.org/10.1145/1978942.1978961
[66] Stuart Russell and Peter Norvig. 2002. Artificial intelligence: a modern approach. (2002).
[67] Jeffrey Saltz, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. 2019. Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)* 19, 4 (2019), 1–26.
[68] Jeffrey S Saltz, Neil I Dewar, and Robert Heckman. 2018. Key concepts for a data science ethics curriculum. In *Proceedings of the 49th ACM technical symposium on computer science education*. 952–957.
[69] Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R Brubaker. 2020. How We've Taught Algorithms to See Identity: Constructing Race and Gender in Image Databases for Facial Analysis. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–35.
[70] Clifford Siskin and William Warner. 2019. Is It Time to Dezone Knowledge?
[71] Sami Surakka. 2005. Specialization in Software Systems: Content analysis of degree requirements. In *Now that the conference is successfully completed and even the financial affairs taken care of, I would like to thank the invited speakers for their talks and the members of the Organizing Committee for their precious contribution in practical arrangements. Also, I would like to thank the 46 participants from 7 countries for being there and making a jolly good conference.*, Vol. 67. 162.
[72] James E Tomayko. 1998. Forging a discipline: An outline history of software engineering education. *Annals of Software Engineering* 6, 1-4 (1998), 3–18.
[73] Stephen Toulmin. 1973. Human understanding, Vol. I: The collective use and evolution of concepts. (1973).
[74] Sherry Turkle and Seymour Papert. 1990. Epistemological pluralism: Styles and voices within the computer culture. *Signs: Journal of women in culture and society* 16, 1 (1990), 128–157.
[75] Tamar Vilner and Ela Zur. 2006. Once she makes it, she is there: gender differences in computer science study. In *Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education*. 227–231.
[76] Kiri Wagstaff. 2012. Machine learning that matters. *arXiv preprint arXiv:1206.4656* (2012).
[77] Alicia Nicki Washington. 2020. When Twice as Good Isn't Enough: The Case for Cultural Competence in Computing. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 213–219.
[78] Sarah Myers West, Meredith Whittaker, and Kate Crawford. 2019. Discriminating systems. *AI Now* (2019).
[79] Richard Whitley et al. 2000. *The intellectual and social organization of the sciences*. Oxford University Press on Demand.
[80] J Wiggins. 2017. The Myth of the Full-Stack Unicorn Developer.
[81] Marty J Wolf, Don Gotterbarn, and Michael S Kirkpatrick. 2019. ACM Code of Ethics: Looking Back and Forging Ahead. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 801–802.
[82] Alice Xiang and Inioluwa Deborah Raji. 2019. On the legal compatibility of fairness definitions. *arXiv preprint arXiv:1912.00761* (2019).
[83] Stuart Zweben and Betsy Bizot. 2018. 2017 CRA Taulbee survey. *Computing Research News* 30, 5 (2018), 1–47.