Teaching Data Ethics:
Foundations and Possibilities from Engineering and Computer Science Ethics Education

Anna Lauren Hoffmann
Katherine Alejandra Cross

April 1, 2021
This report is based upon work supported by the National Science Foundation under grant no. SES-1835161.

Table of Contents

Preface 1

About the Authors 2
About the Project 2
Acknowledgements 2

Insights for Teaching Data Ethics from Engineering and Computer Science Ethics Education 3

Introduction 3
Methods and Limitations 4

Justifying Ethics Education: History, Foundations, and Overviews 7
Reference List 9

In The Classroom: Approaches, Designs, and Evaluation 11
Reference List 12

Teaching Strategies: Case Studies, Games, and Role-Playing 17
Reference List 19

Studying Ethics Education: Assessment In and Beyond the Classroom 22
Reference List 23

From Ethics to Virtue and Justice: Critical Interventions 26
Reference List 27

Towards Data Science Ethics Education: Ethics Within and Beyond the “Integration” Debate 31

Reference List 33

Notes 36
Ethical, social, and other considerations have been acknowledged as important for data science education. However, research is only beginning to emerge on how such considerations are being incorporated into data science curricula. The NSF project “Emerging Cultures of Data Science Ethics in the Academy and Industry” has been examining the state, structure, and substance of "data ethics" education in both higher education and industry.

Our work joins other projects discussing the nature and scope of “tech ethics” syllabi (many of them relevant to data science), illuminated the challenges and perspectives of “ethics owners” in corporate contexts, and that have proposed experimental or speculative redesigns of tertiary-level ethics classes. We are motivated by the conversations these and other projects have sparked. They are integral to cultivating curricular interventions that enhance the critical and ethical sensibilities of students and practitioners in data science and technology.

Unfortunately, our initial project plans were upended by covid-19. We are hardly alone in this regard; the pandemic has slowed research, interrupted teaching, and complicated many scholars’ careers for more than a year now. The impact has had a disproportionate impact on certain groups, including (but not limited to) those with young children or other caretaking responsibilities. This has been true for many of us on this project. Nonetheless, our work continues. If anything, the pandemic response’s reliance on numbers, statistics, and data dashboards has made critical and ethical education in data science more urgent.

As an early output of this project, this document reflects an attempt to systematically trace connections and disjunctions between data science ethics education and its precursors in engineering and computer science ethics education. We developed this document to make our work open and accessible beyond the confines of a single project. We hope this work will be valuable to others engaging engineering and computer science ethics education as a departure point for designing data science ethics programming. We hope our framing of relevant issues, frameworks, and strategies will be useful for moving beyond past limitations, uncovering new possibilities, and charting a path for future work.

-Anna & Katherine
About the Authors

Anna Lauren Hoffmann is an assistant professor at The Information School of the University of Washington and affiliate faculty with the UW DataLab. She is a PI on the NSF-funded project “Emerging Cultures of Data Science Ethics in the Academy and Industry.” Her writing on data, technology, and ethics has been published in New Media & Society, Information, Communication, & Society, The Library Quarterly, and the Journal of the Association for Information Science and Technology. Her public scholarship has appeared in various places, including The Guardian, Slate, and The Los Angeles Review of Books. She has been teaching on the social and ethical information of data, information, and technology since 2010.

Katherine Cross is a PhD student at The Information School of the University of Washington. She served as a research assistant on the NSF-funded project “Emerging Cultures of Data Science Ethics in the Academy and Industry.” Her research on ethics, games, and online culture has been published in the NSF-funded project “Diversifying Barbie and Mortal Kombat,” as well as Women’s Studies Quarterly, and Human Technology. Her social and cultural criticism has also been published in venues like Rolling Stone, Slate, The Verge, and The Guardian.

About the Project

The multi-institutional, National Science Foundation-funded project is titled “Emerging Cultures of Data Science Ethics in the Academy and Industry” (SES-1835161) and draws together a diverse group of information ethics and policy researchers to (1) document and assess barriers and opportunities for integrating ethics into data science practice; (2) assess the continuities and discontinuities emerging between industrial and academic contexts; and (3) develop a foundation for cohesive, comprehensive, integrative data ethics education. The investigators will use a range of complementary methods to do so, including qualitative interviews, expert judgement, and quantitative computational analyses of the latent thematic and topical structure of documents from academia and industry.

Project PIs: Deirdre Mulligan (UC Berkeley); Karen Levy (Cornell University); Solon Barocas (Cornell University); Anna Lauren Hoffmann (University of Washington)

Acknowledgements

The authors would like to thank Deirdre, Karen, and Solon for helpful conversations that informed our work. We would also like to thank Carole Palmer (University of Washington) and Megan Finn (University of Washington) for additional input and guidance.
Insights for Teaching Data Ethics from Engineering and Computer Science Ethics Education

Introduction

Data science and technology has established itself as an active field of work and study, especially in higher education. The desire to exploit accumulations of digital data and advanced computational techniques for social, political, or economic gain has found taken hold in a range of domains—from science and medicine to retail and manufacturing to news and entertainment. Accordingly, a wide array of researchers, educators, and practitioners alike have championed the supposedly transformative potential of “big data,” machine learning, and artificial intelligence.

This desire has fueled demand for professional data scientists and related information professionals that employ skillsets endemic to a range of jobs, including (but not limited to) software development, quantitative research, and statistical analysis. Colleges and universities have capitalized on this demand by developing data science curricula, degree programs, and even standalone data science schools and departments within academic institutions. These efforts seek to equip interested students with the skills and competencies considered necessary for such information professionals, offering coursework in statistics and computer programming.

At the same time, a number of factors have raised public, professional, and regulatory awareness of the harmful and even violent nature of data science and technology. Numerous privacy scandals, high profile data breaches, and concerns over things like misinformation, social media manipulation, and algorithmic discrimination have helped foreground harms enabled by “big data,” automated decision systems, and AI. These harms have been made visible and further contextualized by the work of activists, advocates, journalists, educators, researchers, and even some data scientists themselves concerned with problems of data ethics and justice. Accordingly, accounts of the potential of data science and technology are now often coupled with a recognition of its capacity for furthering problems of inequality, oppression, and exploitation.

This tension—between the promises and pitfalls of data science—has both helped raise awareness of existing and stimulated new commitments to ethics in data science and related curricula. Sometimes these commitments are articulated through the lenses of existing, but perhaps limited, applied ethical domains—like engineering ethics, computer science ethics, research ethics, professional ethics and codes of conduct, or individual ethical responsibility. In other cases, it has sparked calls for new commitments, especially critical or progressive engagements with questions of race, gender, ability, and other axes of social organization and control.
Against this uncertain backdrop, a number of academic institutions have attempted to integrate ethics into their data science curricula. A number of universities have recently launched courses on the ethical, legal, and social implications of working with data in their data science curricula, and have begun hiring postdocs and faculty with focused expertise in data ethics. These initiatives aim to establish ethics as a foundational component of data science education and to cultivate disciplinary norms that involve careful ethical deliberation, but they do so without the benefit of a widely shared understanding of what such an education should entail. Meanwhile, efforts to further develop and understand the existing terrain of data ethics education is complicated by institutional challenges, including logistical difficulties induced by adding courses to program curricula, questions about what types of scholars are best positioned to teach data ethics (given its combination of technical, social, and philosophical dimensions), and resistance from faculty who view social and ethical dimensions of STEM fields as tangential.

Given the plurality of social, political, and institutional visions being brought to bear on data science and technology, there is no standard curriculum or agreed-upon set of core data ethics problems or principles. Nor is there an established taxonomy of the issues, challenges, and dilemmas surrounding ethics education, leaving instructors or administrators to wonder: What core normative issues must a data science education address? Where and when do these issues arise? What methods and (inter)disciplinary approaches are best suited to addressing them? And how are they made tractable to data scientists in the course of their professional practice?

As one source of insight into these questions, we reviewed data ethics precursors in engineering and computer science ethics education. There are, of course, other relevant histories. But the relevance of those histories does not diminish the importance of engineering and computer science to the development of contemporary data science and related domains like machine learning and artificial intelligence. The reviews presented here add historical and pedagogical context to a growing body of research on ethics in data science education. Where applicable, we have identified relevant connections and disjunctions to data science ethics education literature as a preliminary sketch of the intellectual terrain.

**Methods and Limitations**

Although ethics is central to domains like engineering and computer science, it has historically occupied a precarious place in the education and training of future professionals. While codes or principles of ethics for the engineering profession have a longer history, systematic and sustained conversations addressing the incorporation of ethics in engineering education do not take hold until after the 1960s. In computing, professional associations—like the Association for Computing Machinery (ACM)—have
at times expressed some ambivalence towards ethics education, as early versions of the ACM’s curriculum recommendations considered societal aspects of computing irrelevant to formal training in computer science. In other ways, however, attitudes appear to have shifted. For example, courses on privacy and security are now commonplace in computer science, though instruction in these areas can sometimes tend toward technical defenses rather than more engaged social, ethical, or values-based inquiry. And as other work has shown, there are perhaps hundreds of college courses that educators and instructors identify as relevant to technology ethics broadly; these courses cover topics ranging from inequality and human rights to civic responsibility and environmental impact. Whatever ambivalence persists, there also exists dedication and even exuberance for ethics in some corners of engineering, computer science, information science, data science, and beyond.

Our high-level review of the literature examines at least a century of writing about the moral, social, and political content of engineering and, later, computer science education. We have organized the literature into five broad categories: (1) Foundations, Theory, and Justifications; (2) Classroom Approaches, Proposals, and Designs; (3) Teaching Exercises and Strategies; (4) Empirical Work and Efficacy; and (5) Topical Approaches and Critical Interventions. This scheme is, no doubt, imperfect. For example, some entries could fit on multiple lists, but we’ve included each only once. Accordingly, our work is not definitive and we do not mean to suggest causal relationships between events or groups. Rather, they should be seen as broad, informed, but ultimately subjective observations on theme, context, and possible connections.

We oriented our search strategies toward work that was explicit about addressing ethics in engineering and computer science education. We grounded our search in ethics and education as topics and search terms, using combinations of “engineering ethics education,” “engineering ethics,” “engineering education,” “engineering ethics pedagogy,” and related terms (e.g., “pedagogy”) to identify relevant articles. We used similar terms for locating work on computer science ethics education. We searched multiple databases, including Academic Search Complete, JSTOR, Web of Science, ACM Digital Library, Springer, ENGnetBASE, and the University of Washington’s general library search portal. We reviewed the results and collected work that directly addressed the topic ethics education in engineering and computer science curricula or courses. We discarded articles that only mentioned ethics education in passing. We also targeted specific journals (e.g., Science and Engineering Ethics, European Journal of Engineering Education, Engineering Studies) for further exploration. Finally, we also combed the bibliographies of more extensive reviews and overviews identified in our initial searches (e.g., Hess and Fore, 2018).
Overall, we collected and summarized themes across 300 works ranging from original research articles in journals and conference proceedings to books to comments and editorials. The bulk of the searching and reviewing efforts were carried out by Katherine (in consultation with Anna) over the summers of 2019 and 2020. The initial results were sorted into more than a dozen ad hoc categories. These categories were subsequently condensed into the five categories contained in this document. In the process, Anna reviewed and synthesized the initial summaries, with Katherine providing further writing and editorial assistance.

Though our efforts turned up a great deal of work, our approach limited us in certain ways. Most obviously, we only consulted English language publications; we can make no claims about history or nature of ethics concerns in engineering and computer science education literature produced in other languages. We also limited our collection to higher education, excluding discussions of teaching ethics in computer science at the primary or secondary education levels. As a matter of genre, we also did not include white papers, agency reports, or corporate documentation in this review (though those works have informed other parts of the project). Rather, our efforts here are focused on published research, comments, reviews, and editorials on engineering and computer science ethics in higher education as found in relevant journals, conference proceedings, and professional or trade publications.

No doubt, we will have missed some relevant work. Our search strategies would not catch, for example, articles discussing the teaching of corporate or social responsibility if those articles did not also mention or connect their work to personal or professional ethics. Our efforts will have also missed focused discussions of particular values (e.g., integrity, professionalism, justice) or pedagogical commitments (e.g., critical pedagogy) if those values or commitments were not connected to “ethics” as a keyword or theme (though our search did identify a number of links to these further conversations, many of which are included in the “Expanding the Curriculum” section).

Despite these limits, we believe our approach provides--at a minimum--a useful departure point for connecting the teaching of data ethics to longer histories of engineering and computer science ethics education. Focusing on “ethics” as a keyword is justifiable given its dominance as an umbrella term for discussing normative considerations in professional contexts. Nonetheless, this document should be taken not as a final authority, but as grounds for further research and exploration.
Justifying Ethics Education: History, Foundations, and Overviews

This list contains items of historical note, snapshots of trends at particular moments, theoretical debates, and some works frequently cited in the research. It is not meant to signal some “core” or canon of engineering ethics education—though some entries may, indeed, be considered “canonical.”

In terms of historical context, our review turned up connections between ethics and the substance of engineering education in the United States at least as early as the 1920s. A 1922 article mentions a disconnect between the different “ideals” and ethics that emerge from engineers with different educational backgrounds, noting that these differences had practical implications for how engineering work was perceived and compensated. Another piece from 1929 describes a range of professional ethical issues confronting the “executive engineer,” declaring that “the educators in our engineering schools should, in designing their courses, bear [these issues] well in mind.”

In a 1940 Science article, MIT’s Dugald C. Jackson describes a sound engineering education as attending to “knowledge of facts of nature and their relationships, and with facts regarding man and his relationships, and to become wise in adapting the phenomena of nature to the service of man.” Notable, if unsurprising, however, this and other early works orient engineering education toward ethics, ideals, and notions of “progress” that promoted (sometimes quite explicitly) white, Euro-centric, and patriarchal superiority and field racist and genocidal projects of discrimination, displacement, and dispossession. For example, Jackson suggests that “inadequate” engineering competence may be responsible for the struggles of Central American and African groups—not the violent impacts of European colonialism and the mid-Atlantic slave trade.

While notable, these early examples are nonetheless few. Starting in the 1970s, however, one sees an expansion in both topics and methodological approaches to engineering ethics education. Texts like Edwin Layton’s 1971 work The Revolt of the Engineers: Social Responsibility and the American Engineering Profession mark a broader shift in thinking about ethics in engineering and, later, computer science education—away from the lofty (and racist) ideals of some forerunners and towards questions of social and political responsibility. Weil’s 1984 paper contextualizes this trend, connecting engineering ethics to the aftermath of Watergate and the social movements of the 1960s. She also illustrates the early use of disaster-based case studies from the start as both justification for and content of ethics classes.

The 1970s and 1980s are also marked by a number of higher-level discussions of the role and implementation of ethics in engineering and computer science education. For example, Deborah Johnson’s germinal “Who Should Teach Computer Ethics and
Computers & Society?" is an exemplar of this period of discussion—and many of the issues captured in this literature find updated and renewed attention in her 2020 text *Engineering Ethics: Contemporary and Enduring debates*. Other work asks: what should we be teaching? And how should we be teaching it? These conversations were further bolstered by the formation of working and interest groups within certain professional associations, like the American Society for Engineering Education. In some sense, then, data science ethics education has merely reinvented the ethical wheel, reproducing ongoing and unsettled debates from prior decades.

From the 1990s onward, two further trends are evident. First, we see a proliferation of empirical work on perceptions of ethics, surveys of existing offerings and efforts, and the efficacy of ethics education. In some ways, this is reflective of the broader rise of “auditing cultures” and neoliberal governance in institutions of higher education, where significant amounts of time and resources were diverted into auditing, evaluation, and the development of quantifiable metrics of “success.” Indeed, some work from this period onward cites events like the 1989 NSF sponsored workshop on the “Introduction of Legal, Ethical, and Values Issues in Engineering Education” as grounding. This workshop and other events articulated explicit calls for more systematic evaluations of ethics in engineering education and other domains.

Second, while concerns from social movements of the middle and late 20th century impacted other parts of the profession, they only begin to clearly register in engineering ethics education research later, notably through the incorporation of feminist ethics and attention to issues of racial and ethnic diversity. This lag perhaps lends some credence to the idea that certain domains of work have been slow to “catch up” in terms of grappling with legacies of racist and sexist exclusion in and out of the classroom. Conversely, one might say that the lag is indicative of the resiliency of white supremacy and patriarchal domination—and a testament to the work of activists, advocates, and others who have confronted extant powers and opened up opportunities for others. At the same time, one should be mindful that just because certain voices or groups are not reflected in certain literatures does not mean, people of color, white women, disabled people, religious minorities, and/or LGBTQ people were absent or were not shaping ethical practice in other ways. Further, practitioners, scholars, and professional organizations were addressing problems of exclusion and representation before the 1990s, but not necessarily under a framework of “ethics.” The gap between these efforts and ethics education research underscores a continuing need for sources and methods that bridge these divides and open up alternative narratives and ethical possibilities.
Reference List


In The Classroom: Approaches, Designs, and Evaluation

This list contains work on the design and content of ethics education in engineering and computer science. Some of the work discusses actual classes that have been developed and, in some cases, actually taught by the authors. Other papers take a more broad and theoretical approach, discussing particular models for course design or the applicability of particular pedagogical theories. Some further works here also address how to evaluate courses or programs, without necessarily rising to the level of a systematic study of the efficacy of a given program or approach.

Some of the work here discusses ethics in engineering and computer science education at the curricular or programmatic level. Some of these papers can be read as helping to confer validity, offering “official” published accounts to help interested faculty justify the need for ethics courses or content to administrators or other professionals. A key example here is “Implementing A Tenth Strand in the CS Curriculum” (Martin et al., 1996) which laid out a systematic vision for adding ethical competencies to the other competencies detailed in the influential Computing Curricula 1991.

Starting around the year 2000, we also begin to see greater push back against approaches to teaching ethics seen as “conventional” or that rely largely on the individualization of responsibility and abstract ethical debate. Some of this push back takes the form of broad calls for project-based coursework or more inter- or multi-disciplinary engagement. Other pieces mobilize around the US National Academy of Engineering’s 2008 “Grand Challenges,” which set forth a series of issues that the Academy believed to be the greatest ethical and technical challenges for the profession in the twenty-first century. Some works cite the “Grand Challenges” as useful for advancing multidisciplinary and cross-unit initiatives on urgent issues like sustainability. Other work is more critical, suggesting that the Grand Challenges are themselves in need of ethical assessment and arguing for greater attention to “the ethics of problem framing, and the consideration of social justice questions as an integral part of professional ethics” (Riley, 2012, p. 123). These pieces provide an important link to the more critical and justice-oriented approaches we detail later (see “From Ethics to Virtue and Justice”).

Other work featured here is less general, focusing instead on specific approaches to course design. Some approaches are comprehensive, seeking to connect theory, pedagogy, course structure, and evaluation. For example, Bairaktarova and Woodcock (2017) lay out an in-depth, theoretical model for teaching that combines the theory of planned behaviour with perceived spheres of control—that is, spheres where an individual feels they have personal agency. Others are more narrowly tailored, focusing
either design or evaluation. On evaluation, for example, Mumford, Steele, and Watts (2015) propose a multilevel approach to evaluating the effectiveness of engineering ethics classes, arguing that “multiple measures can and should be employed” including “measures examine behavior, cognition, reactions, and institutional outcomes.” Combined, these works model systematic thinking about ethics education in the classroom, from concept to implementation.

On the question of course design, this section also features debates on how best to integrate ethical considerations into a curriculum. This debate is often, though not always, framed as a choice between substantive standalone ethics courses versus weaving ethical issues throughout the program, for example in introductory programming classes or in machine learning courses. Saltz et al. (2019), for example, argue “that ethics content should be integrated into core computer science classes, as a preferable solution over simply having a standalone ethics class,” adding that “ethics should be seen as a necessary part of daily practice rather than a public relations digression from what is actually important” (p. 32:3). Other integrative approaches include the modular Harvard EthiCS program, where working a module into each of fourteen different classes with each module was taught by a graduate student with expertise in the relevant area (often from the Philosophy department) rather than the course’s instructor of record. As we will discuss in the closing section (“Towards Data Science Ethics Education”) these debates have had significant overlap with early discussions of how to incorporate ethics into data science education.

Reference List
Bernstein, M. J., Reifschneider, K., Bennett, I., & Wetmore, J. M. (2017). Science Outside the Lab: Helping Graduate Students in Science and Engineering Understand the Complexities of Science


Teaching Strategies: Case Studies, Games, and Role-Playing

One of the most significant continuities between the engineering literature and the data science literature is the pedagogical use of case analysis—that is, the analysis of case studies relevant to the course material. They are configured as vital to the enterprise. Case studies encompass everything from historical events to fictionalised accounts inspired by real events, and each often focuses on dilemmas: difficult choices between competing and compelling interests. Speaking very broadly, these choices are framed as individual decisions. While not every case study will have these elements, model case studies—especially fictionalised case studies built from the ground up—often will. In every event, case analysis offers the ethicist empirical data for presentation and study. It gives their students something to ‘play’ with that affords them a sense of the real-world dimensions of their jobs.

Some case-based strategies take a more performative approach, asking students to act out roles and specified scenarios—some fiction, some grounded in real events. For example, Dodson et.al. (2019) who use a roleplaying exercise to put students in the roles of various stakeholders in a 19th century industrial city. Meanwhile, Wilson (2013) describes an exercise where his students “took on the role of a faction involved in the Chernobyl disaster,” an approach that takes up the case study approach but modifies it slightly by having students to not only read and debate a case but act out its details. Similarly, Monk (2009) advocates using plays as old as Antigone to teach engineering ethics because of its themes about engaging with, and challenging authority. He regards the character Creon’s “preoccupation with the smooth running of the machinery of the city” as something “that suggests that the play may be relevant to engineers” (p. 115).

Case study methods have been further critiqued as positioning ethics as a kind of detached and individual endeavor. Lynch and Kline (2000), for example, argue against cases that model what they called “brinksmanship ethics”—that is, a focus on debates and scenarios that emphasize heroic individual whistle-blowers and ignore broader “sociological and cultural context” (p. 208). One finds similar critiques in, for example, medical ethics, where an overemphasis on “difficult and tragic cases” are found to be “inadequate for developing the moral agency of the student” (Liaschenko, Orguz, and Brunnquell, 2006, p. 675). Here, case studies are said to abstract the student away from the reality of the scenario they face by situating them as experts on the privileged end of a power relationship (teacher-student, provider-patient). Other approaches might instead emphasize collaboration, with time set aside “to look specifically at relationships and institutional and social contexts,” with the aim of helping students become “ethics consultants rather than ethics experts” (Liaschenko, Orguz, and Brunnquell, 2006, p. 675).
Other classroom strategies involve further appeals to art and literature—in particular science fiction—to activate and engage students’ ethical sensibilities. Bates et. al. (2012) offers a brief bibliography that illustrates the popularity of sci-fi in CS ethics classes, with examples ranging from the obvious (Asimov’s ‘Laws of Robotics’) to less well known sci-fi and sci-fi films. Burton, Goldsmith, and Mattei (2016) adopt a slightly different approach wherein they use an early sci-fi story, “The Machine Stops” written by E.M. Forester in 1909, to teach computer ethics. The purpose, they argue, is to tease out themes of isolation, dependency, and remote availability without invoking current technologies to which students may have strong attachments.

In some cases, sci-fi is positioned as somehow ‘safer’ than case studies based on real events, as it allows for the simulation of a case study that transcends the limitations (and norms) of reality in order to allow students to explore a situation and apply what they’ve learned more freely. Through examining imaginative examples of storytelling about technology, students can cultivate the kind of imagination in themselves necessary to respond sensitively and thoughtfully to unique cases that may arise in their professional lives. For example, Burton, Goldsmith, and Mattei (2018) argue, “using fiction to teach ethics allows students to safely discuss and reason about difficult and emotionally charged issues without making the discussion personal.” For them, fiction offers both “immersion and distance.”

Finally, one also finds encounters many examples of games being used to teach both engineering and computer science ethics. Dyrud (1998), for example, is quite optimistic about games’ utility for ethics education. In particular, she centers the Lockheed-Martin Ethics Challenge Game—an ethics training board game developed in the 1990s—which shows up in other research as well (see: Bairaktarova and Woodcock, 2015). Heymann and Greef (2018) report their experiences using “serious games”—i.e. games that have some serious, self-consciously didactic purpose—as capstone projects for undergraduates. The design spec proposed attaching a device to children and transmit the data gathered to a computer game. Heymann and Greef argued this project exposed students to problems that went beyond the technical, which included ethical issues around researching children, data use, and education. In a different vein, Gehringer and Peddycord (2013) demonstrate the use of game logic to teach ethics, in this case gamifying the course through the use of “experience points” awarded throughout the term, à la Dungeons & Dragons-style roleplaying game. Teaching through games (see: Heyman and Greeff, 2018; Dyrud, 1998) here evolves into teaching as a game.
Notably, however, this and other games often conflate different levels of ethical analysis. The Lockheed-Martin Ethics Challenge Game, for example, places HR issues alongside larger political or even existential questions—for example, a question about whether or not you report a coworker to management for smoking marijuana in the warehouse mingles with dilemmas about bribery by foreign governments. Though perhaps reflective of the different levels of ethical engagement some professionals might face, it is also indicative of ethics being, as Emanuel Moss and Jake Metcalf have put it, “too big a word”—they note that in data science contexts it is often stretched to account for a wide range of sometimes conflicting processes, outcomes, and values.26

Reference List


Studying Ethics Education: Assessment In and Beyond the Classroom

This reading list covers the broad base of ‘real-world’ studies aimed at understanding the efficacy or impact of particular teaching techniques or lesson plans, inclusion of specific subject matter, or curricular interventions. Some research in this area focuses on teaching interventions or modules that are tested and evaluated in actual university courses and classrooms (e.g., Feldhaus and Fox, 2004); others offer empirical assessments of a range of existing courses, syllabi, or modules (e.g., Bielefeldt et al., 2017). Analysis of outcomes is especially prominent, and generally achieved through surveys—many of which were longitudinal, tracking a cohort of students’ progress through an undergraduate degree programme. For instance, Erin Cech’s (2014) “Culture of Disengagement” paper seeks to understanding flaws in engineering ethics education, concluding from her survey work at four universities that engineering students’ “public welfare concerns” may have actually suffered a decline over the course of their training.

Cech’s work is characteristic of how the survey methodology is often applied, and it points to a specific concern that surveys are often used to address: are students learning to be socially responsible or not? (See, for another example: Hashemian and Loui, 2010). Other empirical work operationalised this question by conducting comparative surveys. One study compared civil, environmental, and mechanical engineering students’ views on social responsibility, finding that students in each had significantly different “positive social responsibility attitudes,” on a spectrum where environmental engineering attracted the most socially responsible students and mechanical engineering the least (Canney and Bielefeldt, 2015). Crucially, environmental engineering students suffered a decline in positive attitudes towards social responsibility over the course of their learning, echoing Cech’s findings. Later work surveys faculty who teach environmental engineering, comparing their responses to those who teach civil, chemical, and mechanical engineering—among other findings, ethical education about poverty and social justice issues were more common in environmental engineering than in other disciplines, but all faculty felt their students received insufficient ethical education. Still other surveys find, for instance, that while computer science educators rate ethical education as a significant concern, many are not formally trained by their universities in how best to teach it (Spradling, Soh, and Ansorge, 2008).

That insufficiency is addressed by the empirical literature in other ways: operationalising and teaching empathy (Walther, Miller, and Sochacka, 2017), for instance, or using novel techniques to better engage students with the material, such as social media or
Many other papers examine a specific course, or a specific problem set and analyse how students dealt with them (e.g., Nudelman and English, 2019). A more comprehensive example can be found in the computer science ethics literature, where accessibility emerges as a concern—particularly in the human-computer interaction (HCI) sub-discipline (e.g., Palan et al., 2017).

Across this list, however, one gets a sense that despite sustained interest or attention, addressing ethics in engineering and computer science engineering education remains a significant challenge. Relatedly, there is some concern in the literature here that engineering and CS ethics education has difficulty reliably producing ethical graduates. The literature does not necessarily reflect heated debate, but rather different researchers and educators offering different lenses on the source of the problem. McGinn (2003), for instance, argues that empirical rigour needs to be brought to defining ethical issues and values in ways that are specific to engineering, while Kert, Uz, and Gecü (2014) identified ways that technology can be used to enhance learning in the computer science ethics classroom. Finally, there remains ongoing discussion on which learning objectives are most desirable in the classroom, and a perceived mismatch between learning approaches and educational goals (e.g., Keefer, Wilson, Dankowicz, and Loui, 2013).

Reference List


From Ethics to Virtue and Justice: Critical Interventions

This list contains works that identify as variously “critical,” insurgent, or as pushing against dominant paradigms of engineering and computer science ethics education—that is, against paradigms that center individual responsibility, atomistic modes of ethical reasoning and agency, or that bracket relevant histories of discrimination and exclusion. These critical responses include proposals grounded in concerns for social justice, feminist or care ethics, inter- and cross-cultural ethics, and narrative ethics. In this way, these works extend and make good on the gestures toward more inter- and multi-disciplinary research signaled in some of the previous lists. By confining these works to a separate list, we do not mean to marginalize them. Rather, we intend this list to be a kind of benchmark for identifying the limits and professional dogmas that mark other works, especially on the foundations and course approaches lists. Collectively, they are useful for unsettling assumptions about what constitutes “ethics” and ethical reasoning for the engineer or computer scientist.

Some entries on this list call for engineering ethics education to be more emotionally involved, more complex, and more responsive to the particulars of day to day life. To this end, some works seek to emphasize the “social” or the “societal” against the individual or atomistic; other work draws on the virtue ethics tradition to highlight the role of character for the ethical engineer or computer scientist. In a slightly different way, Newberry (2004) centers “emotional engagement,” seeking to cultivate students’ emotional investment in ethics and their sense that it is useful to their careers. This work inveighs against the “engineer-ization” of ethics and the reduction of complex ethical quandaries to mere design problems (p. 350). For this and other work, engineers must be taught to think “in non-engineering ways” and resist the temptation to see every problem as reducible to a technical solution.

Other interventions seek to draw explicit connections between different disciplinary or cultural approaches to normative concerns in engineering and computer science. Wang, Zhang, and Zhu (2015), for example, lay out a framework that draws on Chinese sociology of engineering, the Dutch School of engineering philosophy, and Confucian philosophy to argue for a unity of theory and practise. It is vital, they argue, for engineering programs to teach students how to judge and interpret unique situations, and how to interpret their own codes of ethics. This work joins others emphasizing the need for interdisciplinary thinking to realize broadly ethical commitments to ideals like sustainability, social responsibility, or environmental stewardship (e.g., Johnson and Wetmore, 2008; Cumming-Potvin and Currie, 2013).
Other research draws explicitly on knowledges and ethics associated with particular social movements or justice- and identity-based concerns. Feminist methods and ethics feature prominently in this regard. Riley (2013), for example, defends feminist care ethics from those who would regard it as essentialist or otherwise reinforcing of women’s subordination; she also pushes back on other work that scrubs care ethics of its feminist content (e.g., Pantazidou and Nair, 1999) and argues that “scholars doing feminist work must be able to use the word openly without negative repercussions.” In a different vein, work by Jordan (2011), Vakil (2018), Jiménez, Pascual, and Mejía (2019), and others calls for a shift from ethics to justice, challenging ethics education to focus less on individual responsibility and more on inequality, historical injustices, diversity and inclusion, and broader social impacts. These works parallel works from the empirical study list that seek to identify and evaluate the experiences of particular underrepresented or historically marginalized groups, in particular the experiences of lesbian, gay, bisexual, and transgender students in engineer (e.g., Cech and Rothwell, 2018).

It is worth noting, however, that our search efforts turned up comparatively little work accounting for or addressing histories of racial and ethnic discrimination, exclusion, and violence (for some exceptions, see: Kant, 2015; Bielefeldt et al., 2018). While feminist perspectives register clearly, there are far fewer engagements with--for example--critical race theory, ethnic studies, and related domains. On the one hand, this might simply mark a limit of our search strategies. On the other hand, it could also signal that such concerns have not gained significant purchase in ethics education research for engineers and computer scientists specifically. If that is the case, then it is a particularly egregious oversight given the role that non-white groups have played in the history of engineering in the United States—from the labor of enslaved Black and African people to Chinese migrants and other immigrants in building the nation’s infrastructure throughout the 19th century, to the contemporary contributions of Black software engineers and south and east Asian immigrant workers in the tech industry, there are countless examples of such contributions, all structured by politics and power and which receive too little attention in ethics education.

Reference List


Engineering Deans Council (WEEF-GEDC), 1–4. https://doi.org/10.1109/WEEF-GEDC.2018.8629655


29


Towards Data Science Ethics Education: Ethics Within and Beyond the “Integration” Debate

This final section represents a selection of work on incorporating ethics into data science education. It reflects an emergent domain still accreting academic coherence and canon, but also one where distinct themes have taken shape, such as the ongoing debate about where ethics belongs in the curriculum. In this, it has much in common with the disciplinary ancestors identified in this report. Given that overlap between some of the competencies in data science and computer science, many of the items in this section could be readily incorporated into the preceding lists. At times, our sorting here was somewhat arbitrary. However, we wanted to use this list to identify articles that highlighted ethical concerns tied directly to data science and technology, even if they at times articulate those concerns through the lens of computer ethics or the more amorphous “tech ethics.”

Mirroring what we saw with the engineering ethics education literature, one prominent debate in data science ethics education literature concerns the merits of integrating ethics across the whole curriculum as opposed to siloing it in a specific class. Integration can mean slightly different things, depending on where one looks, but it tends to refer to a holistic approach that weaves ethics education into most or all of the courses in a given major. In the engineering literature there is sometimes a distinction made between ethics education being “tacked on” to the end of a course, or isolated in a specific unit amidst a larger course, and ethics being more seamlessly woven into the course material at every stage (e.g. Hirsch et. al. 2005; Davis 2006). In that literature, “integration” often refers to the latter approach.

In emerging work on data science ethics education, there is widespread support for integrative approaches. For some, integrating ethics across courses or a curriculum is more reflective of the way data scientists and other information professionals actually encounter ethical issues--that is, ethical issues are not separate or subsequent to data science, but arise during the course of doing data scientific work. As Baumer et al. (2020) put it, “our students are part of the generation of data scientists that will address these issues and restore faith in data-driven applications. In order to do this, they need to see weighing ethical considerations as an integral part of the process of doing data science” (p. 25). This parallels sentiments expressed in engineering ethics education literature, as when Moore et al. (2005) claim that “unless we can get students to see ethical decision-making as a ‘routine part of engineering’ we will never be successful in teaching the topics that lead to that skill” (p. 274). Interest in integrating ethics, then, is
hardly new, but is a persistent theme across decades of work on ethics education in engineering, computer science, and data science.

Despite decades of word, however, there is still widespread divergence in how to integrate ethics into courses or curricula. A number of approaches adopt a “modular” approach, in a similar vein as the Harvard EthiCS model discussed earlier. Modular approaches seek to develop more or less transferable modules covering relevant technical or methods skills and the ethical issues they might entail. Baumer et al. (2020) advocate for such an approach, developing an “inquiry framework” that prompts students to ask critical ethical questions during modules on, for example, machine learning processes. Sample questions include: “how do you know the data is ethically available for its intended use?” or “how might individuals’ privacy and anonymity be impinged via aggregation and linking of the data?” Presumably, such an inquiry-based orientation could be incorporated more or less seamlessly into a range of modules. At the same time, however, leading by inquiry does not necessarily address the problem of equipping students with the frameworks or knowledge necessary to address or respond to ethical challenges.

Other integrative approaches forgo modules in favor of more holistic thinking. For example, Stodden’s “Data Science Life Cycle” (2020) approach seeks to address a range of issues across “the complete bundle of artifacts…and knowledge (scientific results) produced in the course of data science research results,” with an aim to get students and researchers to think “beyond the dataset” (p. 61-2). She proposes using the Life Cycle to create a taxonomy of courses in data science programmes, arguing “it helps define a curriculum by using the steps of the Data Science Life Cycle as a pedagogical sequence and provides for the inclusion of overarching topics such as data science ethics, and intellectual property, reproducibility, or data governance considerations” (p. 65).

Despite enthusiasm for integrative approaches, standalone ethics courses also continue to garner interest. Stodden (2020), for example, still calls for the creation of “Data Science Ethics” classes as standalone enterprises, despite otherwise advocating for other kinds of more holistic, integrative thinking. Standalone courses benefit from being able to address ethical, social, and political issues in more detail. In an analysis of ethics in AI education, for example, Garrett, Beard, and Fiesler (2020) compared standalone AI ethics courses to technical AI courses that sought to address ethics and that only the standalone courses seemed to explore the social context of AI in great detail, with particular emphasis on topics like criminal justice, the military, and healthcare (compared to “integrative” technical courses which appeared to focus mostly on narrower problems of privacy or technical framings of algorithmic bias).
A cursory review of engineering and computer ethics education literature shows that there are a range of works to draw on for thinking further about standalone versus integrative approaches. However, there are perhaps further considerations for data science ethics education that are less well developed in these precursors. These challenges will, it seems, need to continue developing new tools and methods for cultivating data science students’ ethical sensibilities. Some of these issues include: engaging a wider range of applied ethical domains, as with research ethics, journalism ethics, or business ethics; incorporating histories of discrimination and civil rights into the curriculum, especially as they relate to algorithmic discrimination and automated decision systems; moving beyond limited or atomistic conceptions of privacy in the face of widespread inference techniques and predictive analytics; or engaging more directly with projects of political refusal and bans on certain technologies, as with debates over facial recognition systems. In this way, data science ethics education has an opportunity to learn from its precursors in engineering and computer science and develop a more expansive terrain of ethical engagement and debate.

Reference List


Notes

[1] The authors would like to thank Deirdre Mulligan, Karen Levy, and Solon Barocas for collective research and writing that informed portions of this introduction.


Importantly, data science not only entails programming and other skills often associated with computing, it also requires competencies in research methods—especially quantitative methods and statistics. Accordingly, data ethics also demands attention to histories of ethics and the development of ethical commitments from other applied and professional domains, like research ethics, biomedical ethics,


[19] For example, there were a number of items that were not substantively about ethics education but mentioned attention to ethics within engineering or computer science education as a possible way to address ethical problems (see, for example: Leugenbiehl, H. C. (1983). Codes of Ethics and the Moral Education of Engineers. *Business and Professional Ethics Journal*, 2(4), 41–61; Crawford-Brown, D. J. (1997). Virtue as the basis of engineering ethics. *Science and Engineering Ethics*, 3(4), 481–489).


