RITPU | IJTHE
Revue internationale des technologies en pédagogie universitaire
International Journal of Technologies in Higher Education
RITPU.CA
IJTHE.CA

ISSN 1708-7570 Volume 22, n° 1, article 18 2025

Ethics Education in Technical AI Programs: Toward a Complexity and Care-Based Approach

Formation éthique dans les programmes techniques d'IA : vers une approche basée sur la complexité et le *care*

https://doi.org/10.18162/ritpu-2025-v22n1-18

Adèle AUBIN ✉ **Concordia University, Canada**

Available online: September 19, 2025

Abstract

Current initiatives to include ethics in artificial intelligence (AI) technical curricula are slow to bear fruit. This paper discusses several key challenges that undermine these pedagogical efforts to include ethics, namely the confusion around the meaning of AI ethics, a tendency to prioritize technical over social perspectives, and a lack of solid educational grounding. These challenges are further explored using humanities-based perspectives. Finally, potential answers based on the paradigm of complexity and the ethics of care are suggested in response to the identified pitfalls.

Keywords

AI ethics, ethics education, technical AI programs, paradigm of complexity, ethics of care

Résumé

Les initiatives actuelles pour inclure l'éthique aux programmes techniques d'intelligence artificielle (IA) tardent à porter leurs fruits. Cet article discute de défis clés qui sapent ces initiatives d'enseignement pour inclure l'éthique, soit le sens vague de l'éthique de l'IA, une tendance à prioriser la perspective technique plutôt que sociale, ainsi que des bases didactiques peu solides. Ces difficultés sont examinées plus en détail à l'aide de perspectives issues des sciences humaines. Enfin, des pistes de solution basées sur le paradigme de complexité et l'éthique du *care* sont suggérées pour répondre aux écueils identifiés.

Mots-clés

Éthique de l'IA, formation éthique, programmes techniques d'IA, paradigme de complexité, éthique du *care*

1. Introduction

The growing popularity of artificial intelligence (AI) is inciting the frenetic deployment of systems whose socio-ethical impact is often worrisome. While AI ethics is attracting more attention recently (Borenstein et al., 2021), this interest remains marginal. Indeed, initiatives to anchor AI in a quest for the common good account for a paltry 2.6% of AI publications (Fosso Wamba et al., 2021). Among the various means evoked to promote and accelerate this integration of ethics in AI, the need to include an ethical dimension in future AI workers' training is often mentioned.

Obviously, this employee-based approach is not a panacea. Singling out AI developers as the best people to deal with this technology's socio-ethical issues is indeed problematic (Stark & Hoffmann, 2019), since, among other things, it elides the responsibility of companies (Green, 2021; Greene et al., 2019) and governments (Cath et al., 2017). It also neglects the more systemic and institutionalized nature of these issues, over which programmers have little control (Greene et al., 2019; Hoffmann, 2019; Metcalf et al., 2019; Raji et al., 2021).

Thus, despite the frequent use of this narrative, the developers' lack of ethics should not be cited as the only possible explanation for the various issues raised by AI (Greene et al., 2019). Nevertheless, AI workers still have some influence on the design of this technology by transmitting their biases and values (Green, 2021; Greene et al., 2019). Hence, the potential of teaching AI ethics is worth considering, as such efforts do yield some results (Bates et al., 2020; Borenstein & Howard, 2021; Hagendorff, 2020; Tuovinen & Rohunen, 2021) and current regulatory measures are slow to produce real change (Gorur et al., 2020).

That said, the road to true ethics education in AI technical programs is far from smooth. The various initiatives in this direction may be laudable, but the sustainability of their results remains to be seen. Thus, in this paper, I argue that a shift based on the paradigm of complexity and the ethics of care is needed for AI ethics education to become meaningful. To do so, I begin by unpacking the three main elements that, according to my literature review, stand in the way of a functional implementation of an AI ethics pedagogy. I then continue by drawing on humanities-based perspectives to outline the main mechanisms responsible for these difficulties. Finally, I show how complexity and care can help address the identified challenges.

2. Three key Challenges for AI Ethics Education

In an attempt to understand the elements that undermine AI ethics education efforts, I present a literature review that highlights three key challenges: (1) the difficulty of determining what AI ethics means, (2) the primacy given to technical approaches, and (3) the absence of a solid educational grounding in the development of these educational approaches.

2.1 A Hard-to-Define Topic

2.1.1 Negotiating the Meaning of Ethics

The proliferation of codes, guides and declarations is evidence of a certain common ground that has developed around AI ethics and the need to tackle this issue to promote a socially acceptable and beneficial technology while minimizing any harmful consequences (Floridi, 2018; Greene et al., 2019; Hickok, 2021; Jobin et al., 2019; Morley et al., 2020; Stark & Hoffmann, 2019). That said, the notion of AI ethics remains blurry in these various documents, which are nonetheless the most popular means to date of proposing ethical frameworks for the development and use of AI (Green, 2021). In their comprehensive review of 84 AI ethics guides, Jobin et al. (2019) effectively

showed that, while their research identified eleven common general principles (e.g. transparency, solidarity), none of these principles was shared by all of the analyzed publications, and only the five most frequent principles were mentioned in more than half of the literature. This proliferation of resources attempting to define ethical principles in AI thus testifies to the complexity of the task, while at the same time creating further confusion by multiplying the number of voices involved (Munn, 2022; Schiff et al., 2021).

What is more, even if we choose to celebrate the degree of common ground that exists on the identification of principles, their genuine meaning is still far from resolved (Cath et al., 2017; Jobin et al., 2019), notably because of the tensions caused by the many stakeholders involved (governments, industry, civic groups, etc.) (Green, 2021; Greene et al., 2019; Mittelstadt, 2019; Munn, 2022). This lack of consensus keeps the definition of AI ethics at a highly theoretical and vague level, complicating the development of frameworks, objectives or evaluation measures to determine whether these ethical principles are truly observed and implemented (Crawford & Calo, 2016; Green, 2021; Hagendorff, 2020; Jobin et al., 2019; Metcalf et al., 2019; Mittelstadt, 2019; Morley et al., 2020; Schiff et al., 2021). Naturally, this ambiguity is also reflected in technical efforts to develop more ethical AI (Borenstein & Howard, 2021; Hagendorff, 2020; Jobin et al., 2019; Mittelstadt, 2019; Munn, 2022). This undermines the ability to find a solution that remains relevant while also accommodating the plurality of viewpoints involved (Corbett-Davies et al., 2017; Hanif et al., 2021; Harrison et al., 2020; Hoffmann, 2019).

2.1.2 Scattered Training

This lack of clarity surrounding the definition of AI ethics is reflected in the choice of topics addressed in courses. In this respect, the variety of topics listed by Fiesler et al. (2020) in their analysis of curricula for courses about the ethics of technology is telling. The 15 identified categories include the following themes: “Law & policy”, “Privacy & surveillance”, “Philosophy”, “Inequality, justice & human rights”, “AI & algorithms”, “Social & environmental impact”, “Civic responsibility & misinformation”, “AI & robots”, “Business & economics”, “Professional ethics”, “Work & labor”, “Design”, “Cybersecurity”, “Research ethics”, and “Medical/health”. Following the classification system outlined in the publication, I categorized the various resources consulted for the present literature review related to the teaching of ethics in AI programs (and computer science more generally).[1] Considering that I could find at least one reference for 13 of the 15 themes identified by Fiesler et al. (2020) and that none of these references included a topic that was not related to these categories, this typology appears credible.

As such, we can see the breadth of topics covered by AI ethics education, reflecting the subject’s broad scope and lack of definition. As a result, training courses run the risk of spreading themselves too thin, or focusing on a few elements while neglecting others that are just as important. For example, most of the references I gathered under the “Social & environmental impact” category dealt with *social* rather than *environmental* impacts. This echoes Garrett et al. (2020) who deplore the lack of interest given to sustainability in AI ethics curricula, although this principle is one of those identified by Jobin et al. (2019). In their review of course outlines in AI ethics, Garett et al. (2020) also regret the omission of issues of AI diversity and accessibility, which are, however, discussed elsewhere in the AI ethics literature. Finally, the heavy focus on laws and norms when dealing with AI ethics is somewhat nonsensical, since the essence of ethics is not legal or

1. The complete results of this exercise can be provided upon request by email.

regulatory, but rather that of a conceptual framework (Green, 2021; Mittelstadt, 2019; Munn, 2022; Rességuier & Rodrigues, 2020; Wagner, 2018).

Thus, professors wishing to integrate AI ethics into their courses often feel caught off guard by the sheer breadth of the subject (Saltz et al., 2019). Moreover, finding the right balance between technical and more theoretical notions to convey the reality of AI's ethical challenges can be a laborious task (Tuovinen & Rohunen, 2021). The thorny and sometimes controversial nature of certain issues can also contribute to the difficulty of teaching this subject, especially with multicultural groups with different sensitivities (Bates et al., 2020).

In short, the abstract and plural complexity of AI ethics is reflected in the various ways it is taught. The variety and specificity of the topics covered bear witness to the challenge faced by the AI community in grasping and transmitting such notions. Current courses in AI ethics therefore run the risk of offering only a very partial or disorganized view of this issue, which can undoubtedly be detrimental.

2.2 The Primacy of Technical Approaches

2.2.1 An Ambient Techno-Solutionism

Despite the glaring complexity of the socio-ethical issues at stake, the current discourse around AI gives precedence to technical over social perspectives. Indeed, Western culture today tends to regard AI as the obvious path to enrichment and progress, neglecting socio-ethical reflection in favour of technological innovation (Hagendorff, 2020). This conviction is echoed, for example, in the various government initiatives that encourage AI training to maintain a technical and financial edge rather than to support contributions to a real societal project (Dutton et al., 2018; Hagendorff, 2020; Morley et al., 2020; Schiff, 2021).

Within companies, ethical considerations are also often considered secondary and too restrictive for profitable and competitive innovation (Hagendorff, 2020; Metcalf et al., 2019; Mittelstadt, 2019). This conception of ethics as a superfluous restriction thus fosters the use of *ethics washing*, i.e., various tactics adopted by the AI industry to give itself the mere appearance of socio-ethical awareness just to evade some additional reproach or regulation (Bietti, 2021; Green, 2021; Hickok, 2021; Jobin et al., 2019; Lauer, 2021; Metcalf et al., 2019; Mittelstadt, 2019; Munn, 2022; Rességuier & Rodrigues, 2020; Stark et al., 2021; Wagner, 2018). This ethical façade reinforces the fantasy that companies can easily respond to socio-ethical issues by proposing technical solutions, thereby deflecting any challenge to this domination that could be mounted by a more social perspective (Green, 2021). In this respect, the idea that technical solutions are portable and independent of the social context in which they are to be implemented is illusorily accommodated (Selbst et al., 2019; Williams et al., 2022).

2.2.2 Educational Impacts

This primacy of technical approaches is reflected in the teaching of AI – and more generally, tech – ethics (Cech, 2014; Metcalf et al., 2019). Courses focus only on the *mastery* of technical skills, rather than on their *exercise* for the common good (Bates et al., 2020; Raji et al., 2021; Tuovinen & Rohunen, 2021). When ethical issues are taught by computer science professors, there is a risk that these issues will be tackled from a technical perspective that oversimplifies them. Social approaches are thus relegated to the background, perpetuating a deficient value system (Bates et al., 2020; Raji et al., 2021). When humanities professors, who have often received more training in ethics, are responsible for teaching these notions, their outsider status can contribute to a

stigmatization of ethics, reinforcing the idea that ethics is external to AI (Bates et al., 2020; Raji et al., 2021). In this way, regardless of the teacher's disciplinary background, the culture of AI programs does not give ethics the consideration it deserves.

This precedence given to the technical approach is also embodied in the format of AI ethics courses. The teaching of ethics in AI programs usually follows one of two approaches: *embedded ethics*, in which ethical concepts are integrated into technical courses, and *stand-alone courses*, which are devoted solely to the subject of AI ethics (Garrett et al., 2020). Embedded approaches only devote a few sessions to the subject (Bogina et al., 2021; Cernadas & Fernández-Delgado, 2021; Furey & Martin, 2018; Horton et al., 2022, 2023; Shen et al., 2021), usually at the end of the semester "if time allows" (Garrett et al., 2020). Moreover, the embedded ethical content is either not evaluated at all or not rigorously evaluated (Bates et al., 2020; Cernadas & Fernández-Delgado, 2021; Horton et al., 2022, 2023). Stand-alone ethics courses guarantee a less superficial study of the topic (Bates et al., 2020; Saltz et al., 2019), but their independent nature reinforces the belief that ethics are disconnected from technical knowledge (Grosz et al., 2019). Moreover, this type of course is typically optional (Bogina et al., 2021; Henderson, 2019; Reich et al., 2020; Tuovinen & Rohunen, 2021; Weiss et al., 2021). Thus, regardless of the format, AI curricula fail to integrate ethics decisively enough to demonstrate that it is a key priority.

Finally, this culture of disengagement can also lead students who care about social impact to give up their studies in engineering and computer science (Carrigan, 2017; Rulifson & Bielefeldt, 2017) or to lose their initial interest in using tech to address social issues (Cech, 2014). This homogenizes the student profile, favouring those who are already adept at a strictly technical vision of AI (Rulifson & Bielefeldt, 2017), and who tend to reduce socio-ethical challenges to simple questions of data and trust (McDonald & Pan, 2020; Williams et al., 2022). For fear of losing students who would disapprove of the replacement of technical elements with allegedly "useless" ethical notions, some programs may stall their efforts to integrate AI ethics (Bates et al., 2020). This creates a vicious circle that sustains social disengagement.

In sum, curricula perpetuate the traditionally techno-solutionist vision of AI, in which all problems are amenable to a strictly technical approach. This approach undermines attempts to include ethics. This culture of social disengagement contributes to a standardization of the personnel who enter the AI field, reducing the chances of debate that a greater diversity of viewpoints could provoke (Crawford & Calo, 2016). This also reinforces the idea that technical specialization is necessary to be able to change the way things are done, and thus better integrate ethics into AI (Greene et al., 2019; Metcalf et al., 2019; Raji et al., 2021).

2.3 A Pedagogical Approach yet to be Defined

2.3.1 Insufficient Objectives

The absence of a solid educational grounding can be observed in the learning objectives targeted by AI ethics courses. In their review of course outlines in technology ethics, Fiesler et al. (2020) report eight typical learning objectives: (1) *critique*, (2) *spot issues*, (3) *make arguments*, (4) *improve communication*, (5) *see multiple perspectives*, (6) *create solutions*, (7) *consider consequences*, and (8) *apply rules.* We can first note that objectives (3) *make arguments* and (4) *improve communication* are more about expression than the development of ethical competence. Objective (6) *create solutions* is highly techno-solutionist which, as discussed, is counterproductive when it comes to discussing AI ethics. Objective (8) *apply rules* relies on the simple implementation of ethical codes, an aptitude which is not, however, synonymous with a

genuinely ethical attitude (McNamara et al., 2018). None of these four objectives actually tackle improving the inclusion of ethics in AI. Objectives (1) *critique*, (2) *spot issues,* (5) *see multiple perspectives* and (7) *consider consequences* are more appropriate as part of an effort to raise awareness of socio-ethical issues. That said, while these are more relevant, raising awareness alone is not enough, because it does not guarantee a more thorough examination of the complex structures at play in AI socio-ethical issues.

In this regard, Borenstein and Howard (2021) remind us of the importance of AI ethics education also cultivating a sense of *professional responsibility*. As previously mentioned, AI workers are not solely responsible for the ethical development of this technology. If we aim to teach AI ethics, however, training cannot be reduced to a simple prism for noting AI's socio-ethical impacts; it must encourage a sense of empathy. But this is a view that has yet to be widely embraced. To my knowledge, only Shapiro et al. (2020) attempt to do this by adopting a philosophy of care.

2.3.2 A Dispersed Implementation

This lack of a robust ethical orientation makes it difficult to put the ambitions of AI ethics education into practice (Garrett et al., 2020; Saltz et al., 2019; Tuovinen & Rohunen, 2021). It is therefore hardly surprising to find a great variety in the approaches tested:

- debates, discussions and essays on case studies or more abstract issues (Bogina et al., 2021; Cernadas & Fernández-Delgado, 2021; Furey & Martin, 2018; Grosz et al., 2019; Henderson, 2019; Reich et al., 2020; Tuovinen & Rohunen, 2021; Weiss et al., 2021),
- role-playing to adopt the perspective of different stakeholders (Horton et al., 2022, 2023; Shapiro et al., 2021; Shen et al., 2021),
- science fiction to explore ethics through imagination (Burton et al., 2018; Weiss et al., 2021),
- analysis grids to assess various ethical issues (Saltz et al., 2019),
- introduction to various ethical theories (Bogina et al., 2021; Burton et al., 2018; Furey & Martin, 2018; Nakatumba-Nabende et al., 2023; Tuovinen & Rohunen, 2021),
- proofs of concept or small technical analysis exercises (Reich et al., 2020; Weiss et al., 2021), and
- a tracking application to reflect on personal data issues (Shapiro et al., 2020).

While many of these interventions show some promise, it is also noteworthy that publications describing AI ethics teaching interventions rarely draw on any theoretical framework. There are a few brief references to various pedagogies, such as experiential learning (Shapiro et al., 2020), active learning (Grosz et al., 2019; Horton et al., 2022, 2023; Shapiro et al., 2021), learning through a deliberation process (Shen et al., 2021) or constructivism (Bogina et al., 2021; Weiss et al., 2021). Often, though, these pedagogies do not seem to guide the intervention in depth. Moreover, except for Bates et al. (2020), who report using autoethnography combined with a provocation-based method, and Shapiro et al. (2020), who use grounded theory, no precise indication of the methodology used to analyze the results was found in the literature.

This dearth, or even absence, of theoretical and methodological foundations hinders the development of a relevant examination of the value of the tested intervention. This is all the more regrettable as many have pointed out that AI ethics and its teaching are eminently more complex topics to teach and assess than the mastery of technical notions (Fiesler et al., 2020; Horton et al.,

2022; Tuovinen & Rohunen, 2021). Even with the aim of raising ethical awareness, some choose to assess unrelated skills, such as the ability to argue (Tuovinen & Rohunen, 2021), or to assess nothing at all (Furey & Martin, 2018). Most studies instead ask students to indicate what they thought of the course or to evaluate their ethical learning or skills (Bogina et al., 2021; Burton et al., 2018; Grosz et al., 2019; Henderson, 2019; Horton et al., 2022, 2023; Reich et al., 2020; Shapiro et al., 2021; Tuovinen & Rohunen, 2021; Weiss et al., 2021). Some pre- and post-intervention comparisons are reported (Bogina et al., 2021; Horton et al., 2022, 2023; Shapiro et al., 2021; Shen et al., 2021), but longitudinal studies on the effects of AI ethics courses are rare (Horton et al., 2023).

In any case, while students seem to appreciate their ethical training to some extent (Grosz et al., 2019; Horton et al., 2022, 2023; Reich et al., 2020; Tuovinen & Rohunen, 2021; Weiss et al., 2021) and claim that it has improved their ethical competence (Bogina et al., 2021; Grosz et al., 2019; Horton et al., 2022, 2023; Shapiro et al., 2021; Shen et al., 2021; Tuovinen & Rohunen, 2021; Weiss et al., 2021), these results must be interpreted with caution. Most of these studies are looking at optional ethics courses. The students who take part are therefore already interested in the subject, and may also believe themselves to be more objective about the course than they truly are (McDonald & Pan, 2020; Williams et al., 2022).

A review article by Hess and Fore (2018) agrees with this perspective, noting that publications dealing with ethics education in technical programs tend only to describe the intervention and report the students' perceptions of it. This lack of coordination and coherence stalls research efforts, since comparisons between the (too) wide variety of approaches and methods are extremely difficult (Fiesler et al., 2020). Even though the vagueness surrounding the learning of AI ethics can support flexibility, it can also discourage teachers, overwhelmed with the many possible directions (Saltz et al., 2019). This difficulty in establishing a dominant paradigm in AI ethics education can be explained by the more social and interpretive epistemological nature of education as a field, as opposed to the more positivist nature of computer science. This may lead to some resistance among people with a technical background to choosing research and evaluation methods that they perceive as insufficiently positivist and overly interpretive (Bates et al., 2020; Grosz et al., 2019; Raji et al., 2021).

3. Contribution of Humanities-Based Perspectives

To provide a more in-depth examination of the factors at the root of these obstacles, I use a humanities-based lens to explain why the aims of AI ethics education are compromised. Then, I suggest that humanities perspectives can also be part of the solution, creating mindset shifts that could concretely help to improve educational initiatives.

3.1 Understanding the Current Situation

3.1.1 A Quest for Simplicity and Objectivity

First, because the scientific method has historically fostered an overwhelming number of scientific and technical advances, the supposedly objective rigour of science and technology has been given a privileged status (Borgmann, 1992). This has encouraged the emergence of a vision in which human beings can tame nature thanks to science and technology, making it more coherent, orderly and functional (Borgmann, 1992; Foucault, 1975). In so doing, science and technology become synonymous with objectivity, efficiency and progress (Ellul, 2008; Slack & Wise, 2015). This perception then allows for the removal of any non-scientific approach to the various decisions

relating to scientific and technological evolution, relying solely on supposedly unbiased data that would reveal the evidence of progress through rigorous measurement (Ellul, 2008; Winner, 2020).

Yet this conception of science and technology as being capable of giving an account of itself, and of reality more generally, obscures much evidence of the far more complex nature of reality. Every theoretical account tends to encompass some phenomena better than others (Kuhn, 2012). For example, as Criado-Perez (2020) and Schwartz Cowan (1985) demonstrate with an abundance of cases, something presumed to be a technological advance has the potential to prove harmful to women if we too naively assume that the male experience is universal. Thus, it is crucial to remember that science and technology are far from neutral, and are always, at least to some extent, the product of different contingent choices we make, whether consciously or not (Braidotti, 2013; Haraway, 1988; Kuhn, 2012; Slack & Wise, 2015).

This faith in technology as capable of achieving simplicity and objectivity may partly explain the difficulties encountered by AI ethics courses. Attempts are made to summarize and measure the various impacts of the complex subject of AI ethics, but these efforts do little to convey the subtleties of reality. This is compounded by the fact that for many in the AI field, technical approaches appear to be a simpler and more objective method of "solving" reality than more social approaches. This tendency toward approaches that appear to be "simple" and "objective" can also be seen in the failure to implement educational approaches that genuinely grapple with AI ethics challenges. The various initiatives in AI ethics education tend to remain rather superficial, and do not sufficiently consider the more complex perspectives proposed by educational and humanities-based approaches.

3.1.2 A Question of Authority

Second, the privilege dynamic created by science and technology is also an important consideration when discussing AI. Any techno-scientific artefact emerges from a particular context and reflects a certain social power structure (Winner, 2020). Therefore, experts are in a dominant position from which non-experts are excluded (Haraway, 1988). By coming to an agreement among themselves, the scientific community imposes its voice on the general population, which has no means of contesting it (Haraway, 1988; Latour, 1987). The latter therefore consents to the introduction of various technical-scientific concepts of which it has only a limited understanding in order to avoid finding itself in an exhausting state of constant vigilance (Latour, 1987). This notion of the expert's authority thus augments the explanation provided by the quest for simplicity and objectivity. Not only are science and technology not neutral, they also perpetuate a relationship of domination in which experts are in a position of power (Foucault, 1975; Winner, 2020). The fact that experts "accept" among themselves the various socio-ethical issues of AI then forces non-experts to do the same, and thereby perpetuates the currently inadequate development of this technology.

The dominant position of science and technology may also shed some light on the challenges faced by AI ethics pedagogy. First, the difficulty of defining AI ethics can be explained by the authority exerted by technical expertise, which imposes methods that are incompatible with the task. Second, approaches to AI ethics are constructed and taught to privilege technical competence and the attendant technical solutions above any other type of contribution, thereby neglecting socio-ethical impacts. Current courses in AI ethics seek to raise awareness of the many socio-ethical issues involved, but do not make students aware of their sometimes unduly privileged authority in these situations. This is all the more damaging since, considering the lack of diversity in AI, the expert point of view generally shows little interest in socio-ethical issues that affect more people from

marginalized groups (Borenstein & Howard, 2021; West et al., 2019). Finally, more social approaches to education seem to be discredited in the context of AI ethics pedagogy, which is strongly influenced by technical traditions.

3.2 Suggesting new Avenues for Reflection

3.2.1 Embrace Complexity

To start with, I propose abandoning this ideal of simplicity and objectivity that influences AI ethics education and instead placing value on the inherent complexity of the topic, echoing Morin's (1990, 2005) paradigm of complexity. This vision questions the current disjunctive and reductionist approaches to understanding reality, acknowledging how uncertainty complicates efforts to make sense of the world and focusing on the dense interconnectedness and interdependence of systems (Morin, 1990, 2005). Additionally, it highlights the fact that a producer of knowledge for a given system is also the product of that same system (Morin, 2005). As such, it challenges the belief in a single absolute truth and instead appreciates the plurality of well-warranted perspectives on reality to better grasp its complexity (Morin, 1990; 2005).

In doing so, we echo Dewey (1959) who emphasizes the importance of diverse perspectives on the world, including both scientific and non-scientific perspectives. This implies accepting that the vision of reality offered by science and technology is partial and situated (Haraway, 1988). Operating on technical principles alone fails to provide the full picture of socio-ethical issues that could be offered by a diverse set of analyses of the different structures involved (Hoffmann, 2019; McDonald & Pan, 2020; Raji et al., 2021). In this way, a more partial, but also more plural, knowledge opens the door to interdisciplinary collaboration and the recognition of a complex reality that changes as exceptions and errors are revealed (Haraway, 1988; Kuhn, 2012; Morin, 1990). More generally, thinking through the implications of developments in science and technology should be connected to questions of social context, to show and criticize their interdependent relationships (Dewey, 1959; Waddington & Weeth Feinstein, 2016).

Teaching AI ethics must therefore be a complex reflexive process, rather than a simple constraint to integrate or a target to hit (Bietti, 2021; Hagendorff, 2020; Rességuier & Rodrigues, 2020; Selbst et al., 2019; Stark & Hoffmann, 2019). We should first accept that AI ethics is itself a complex object. In this way, we aim to highlight this particularity, rather than trying to reduce ethics to more elementary subunits such as simply focusing on policy or misinformation. Second, we also need to *welcome* (and not merely acknowledge) a greater diversity of viewpoints, including those from the humanities, in thinking about AI ethics and the teaching of it (Bietti, 2021; Borenstein et al., 2021; Cath et al., 2017; Crawford & Calo, 2016; Gebru et al., 2021; Jobin et al., 2019; McDonald & Pan, 2020; Mittelstadt, 2019; Morley et al., 2020; Munn, 2022; Raji et al., 2021; Selbst et al., 2019; Stark et al., 2021; Wagner, 2018). Acknowledging the complexity of the topic and the necessity of tackling it from multiple perspectives may well make the search for any consensus or solution more arduous. However, this would seem to be more constructive than the current trend towards oversimplification and/or a merely technical approach, which are definitely insufficient, if not dysfunctional.

3.2.2 Encourage Care

For the validity of complexity to be genuinely acknowledged and dealt with in teaching AI ethics, I suggest that the ethics of care is an essential framing concept, as it is a key tool for questioning the often unwarranted authority that comes with the privilege of expertise. Technical credibility is

undoubtedly necessary, but remains insufficient without a certain amount of empathy (Stark & Hoffmann, 2019). Building on the rejection of a purely "rational" morality, the ethics of care places greater value on solicitude, benevolence and attention to interpersonal relationships (Gilligan, 1982). Care also encourages us to question the ideas of productivity and profit associated with techno-scientific innovation (Puig de la Bellacasa, 2017). Some versions of care ethics emphasize a slower temporality, more in tune with the demands of care, to innovate with benevolence and an authentic recognition of others (Puig de la Bellacasa, 2017).

By adopting the ethics of care, the aim is thus to propose an alternative to the moral distancing to which the ethics of AI and its teaching often lend themselves (Villegas-Galaviz & Martin, 2023). Future AI workers need to understand that their expertise comes with the responsibility to attend to others' needs (Villegas-Galaviz & Martin, 2023). To achieve this, curricula should teach how to care for others, multiplying contacts with others to better understand, help and be concerned about them, all the more given the primary need care represents (Noddings, 2005).

Clearly, the ethics of care cannot be a universal solution for the challenges of AI ethics pedagogy. That said, it allows us to move beyond simplistic approaches by offering a more flexible framework that seeks to take genuine account of the other, their vulnerabilities, context and interdependent relationships (Villegas-Galaviz & Martin, 2023). This care framework also reorients the purpose of AI, its ethics and teaching, to make it a quest for benevolence and inclusion of the other, rather than a primarily technical performance. Of course, this task of cultivating empathy and benevolence through care is a laborious one. Yet, this vision is more plausible than it might first seem, since professions with relatively equivalent power to that of engineers (and developers), such as doctors, are already much more aware of their expertise's responsibility and duty of care (El-Zein & Hedemann, 2016).

3.2.3 Possible Concrete Solutions

To conclude this second part, I present a few avenues through which to explore the potential of including complexity and care in AI ethics education. These remain basic outlines, as the aim of this exercise is simply to identify some practical ways in which these notions can be integrated into teaching.

First, there is substantial potential in approaches based on problem-solving through service-learning projects. Combining learning goals with volunteer work, service learning is an experiential pedagogy which helps develop knowledge and critical reflection through work with communities (of various kinds) to meet their needs (Felten & Clayton, 2011). As argued by El-Zein and Hedemann (2016) and Raji et al. (2021), social issues, rather than technical specialization, could serve as the foundation of engineering (and computer science) curricula. In this way, service learning helps highlight the complexity of socio-ethical problems (Borenstein et al., 2021; Borenstein & Howard, 2021; Naphan-Kingery et al., 2019; Raji et al., 2021; Williams et al., 2022) and fosters the development of moral reasoning in students (Boss, 1994; Lies et al., 2012) while not compromising the teaching of technical content and skills (Novak et al., 2007). Moreover, this approach could be interesting for students who seek to use their technical careers to serve the common good but do not feel the current curricula can help them achieve this goal (Carrigan, 2017; Rulifson & Bielefeldt, 2017).

Second, to further emphasize the complexity of AI ethics and the importance of caring, I suggest using the embedded ethics strategy, in which ethics is integrated into technical courses. Unlike the stand-alone ethics course, this format effectively signals the breadth of socio-ethical issues and the

need to make them an ongoing concern when performing technical tasks (Borenstein & Howard, 2021; Garrett et al., 2020; Raji et al., 2021). To overcome the current shortcomings of this approach, it is essential to include ethical content throughout the semester and across several courses, since these notions should be frequently addressed to provide for sufficient understanding and integration (Horton et al., 2023).

Third, it is necessary to promote interdisciplinary teaching to ensure a proper representation of the multiplicity of legitimate perspectives (Bates et al., 2020; Borenstein et al., 2021; Borenstein & Howard, 2021; Garrett et al., 2020; Raji et al., 2021; Shen et al., 2021; Williams et al., 2022). A multidisciplinary teaching team is an attainable ideal (Furey & Martin, 2018; Grosz et al., 2019; Horton et al., 2022, 2023; Reich et al., 2020), although more costly in terms of time, financial and human resources (Fiesler et al., 2020; Grosz et al., 2019; Horton et al., 2023). A simpler alternative—the option of adding staff trained in other disciplines to computer science departments—can also be considered (Bates et al., 2020; Fiesler et al., 2020; Grosz et al., 2019). Moreover, interdisciplinarity can also be reflected in a pedagogy based on authentic and complex societal problems, as this encourages a search for the best possible solution that goes beyond the current focus on tech-centred approaches (McDonald & Pan, 2020; Raji et al., 2021; Schiff et al., 2021; Williams et al., 2022).

4. Conclusion

In short, AI ethics pedagogy undoubtedly represents a thorny but crucial issue to address if we are to help future AI developers prepare for the socio-ethical issues they will encounter in their work. As such, this paper began by identifying the main obstacles facing AI ethics curricula, based on a literature review. There were three key obstacles: first, determining the meaning and implications of AI ethics is a fundamentally difficult task; second, the technical perspective in AI continues to be generally favoured over the social one; and finally, current AI ethics education lacks of a solid pedagogical foundation to guide the ways in which the socio-ethical issues are addressed.

I have also argued that perspectives from the humanities offer useful explanations for the challenges faced by the pedagogy of AI ethics. These explanations are organized around two axes: the quest for simplicity and objectivity that drives the AI field, and the relationship of authority that places technical experts in a relationship of domination over non-experts. This discussion, grounded in these perspectives, has offered various avenues for reflection that could improve the teaching of AI ethics. Drawing on the paradigm of complexity and the ethics of care, which can provide a broader response to the obstacles encountered by this type of education, some more concrete avenues for solutions were suggested.

While including complexity and care is far from easy, these two concepts have the potential to bring about change in AI ethics education. Indeed, until now, AI ethics has not been given the priority it deserves in technical programs (and in industry, more generally). In other words, simply raising awareness of the socio-ethical issues raised by AI is no longer enough. Future AI workers now need to be taught to recognize the complexity of these issues, as well as to genuinely care about those who may suffer from the increasing integration of AI. Only when this transformation has taken place will AI ethics education become truly relevant and authentic. As such, greater openness to disciplinary and community collaborations represents a first step in the right direction.

Notes

Data Availability

No data supporting the article were collected during the research.

Acknowledgements

This work was supported by the Fonds de recherche du Québec (**https://doi.org/10.69777/341903**). The author also wants to thank David Waddington for his careful reading and helpful feedback.

References

Bates, J., Cameron, D., Checco, A., Clough, P., Hopfgartner, F., Mazumdar, S., Sbaffi, L., Stordy, P., & de la Vega de León, A. (2020). Integrating FATE/critical data studies into data science curricula: Where are we going and how do we get there? In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 425-435). ACM. **https://doi.org/pt6z**

Bietti, E. (2021). From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 210-219). ACM. **https://doi.org/gt9hx9**

Bogina, V., Hartman, A., Kuflik, T., & Shulner-Tal, A. (2021). Educating software and AI stakeholders about algorithmic fairness, accountability, transparency and ethics. *International Journal of Artificial Intelligence in Education*, *32*, 808-833. **https://doi.org/pt62**

Borenstein, J., Grodzinsky, F. S., Howard, A., Miller, K. W., & Wolf, M. J. (2021). AI ethics: A long history and a recent burst of attention. *Computer*, *54*(1), 96-102. **https://doi.org/10.1109/MC.2020.3034950**

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, *1*(1), 61-65. **https://doi.org/ghzjs8**

Borgmann, A. (1992). *Crossing the postmodern divide*. University of Chicago Press.

Boss, J. A. (1994). The effect of community service work on the moral development of college ethics students. *Journal of Moral Education*, *23*(2), 183-198. **https://doi.org/fhdk5x**

Braidotti, R. (2013). *The posthuman*. Polity Press.

Burton, E., Goldsmith, J., & Mattei, N. (2018). How to teach computer ethics through science fiction. *Communications of the ACM*, *61*(8), 54-64. **https://doi.org/10.1145/3154485**

Carrigan, C. M. (2017). Yearning to give back: Searching for social purpose in computer science and engineering. *Frontiers in Psychology*, *8*, Article 1178. **https://doi.org/10.3389/fpsyg.2017.01178**

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2017). Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics*, *24*, 505-528. **https://doi.org/ggj628**

Cech, E. A. (2014). Culture of disengagement in engineering education? *Science, Technology, & Human Values*, *39*(1), 42-72. **https://doi.org/gjjvgv**

Cernadas, E., & Fernández-Delgado, M. (2021). Embedded ethics to teach machine learning courses: An experience. In A. García-Holgado, F. J. García-Peñalvo, C. S. González-González, A. Infante-Moro, & J. C. Infante-Moro (Eds.), *Proceedings of 2021 XI International Conference on Virtual Campus (JICV)* (pp. 91-94). IEEE. **https://doi.org/g693c9**

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 797-806). ACM. **https://doi.org/gf2q38**

Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, *538*(7625), 311-313. **https://doi.org/10.1038/538311a**

Criado-Perez, C. (2020). *Invisible women: Data bias in a world designed for men*. Abrams.

Dewey, J. (1959). *Dewey on education: Selections* (M. S. Dworkin, Ed.). Teachers College Press.

Dutton, T., Barron, B., & Boskovic, G. (2018). *Report on national and regional AI strategies*. Canadian Institute for Advanced Research. **https://cifar.ca/...**

Ellul, J. (2008). *La technique, ou, L'enjeu du siècle* [*The technological society*] (2nd ed.). Economica.

El-Zein, A. H., & Hedemann, C. (2016). Beyond problem solving: Engineering and the public good in the 21st century. *Journal of Cleaner Production*, *137*, 692-700. **https://doi.org/10.1016/j.jclepro.2016.07.129**

Felten, P., & Clayton, P. H. (2011). Service-learning. *New Directions for Teaching and Learning*, *2011*(128), 75-84. **https://doi.org/10.1002/tl.470**

Fiesler, C., Garrett, N., & Beard, N. (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* (pp. 289-295). ACM. **https://doi.org/ggqzkz**

Floridi, L. (2018). Soft ethics and the governance of the digital. *Philosophy & Technology*, *31*(1), 1-8. **https://doi.org/gcz7zt**

Fosso Wamba, S., Bawack, R. E., Guthrie, C., Queiroz, M. M., & Carillo, K. D. A. (2021). Are we preparing for a good AI society? A bibliometric review and research agenda. *Technological Forecasting and Social Change*, *164*, Article 120482. **https://doi.org/10.1016/j.techfore.2020.120482**

Foucault, M. (1975). *Surveiller et punir* [*Discipline and punish*]. Gallimard.

Furey, H., & Martin, F. (2018). AI education matters: A modular approach to AI ethics education. *AI Matters*, *4*(4), 13-15. **https://doi.org/g5vq39**

Garrett, N., Beard, N., & Fiesler, C. (2020). More than "If time allows": The role of ethics in AI education. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 272-278). ACM. **https://doi.org/ghj6jf**

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86-92. **https://doi.org/10.1145/3458723**

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Harvard University Press.

Gorur, R., Hoon, L., & Kowal, E. (2020). Computer science ethics education in Australia – A work in progress. In H. Mitsuhara, Y. Goda, Y. Ohashi, Ma. M. T. Rodrigo, J. Shen, N. Venkatarayalu, G. Wong, M. Yamada, & C.-U. Lei (Eds.), *Proceedings of 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 945-947). IEEE. **https://doi.org/pt63**

Green, B. (2021). The contestation of tech ethics: A sociotechnical approach to technology ethics in practice. *Journal of Social Computing*, *2*(3), 209-225. **https://doi.org/10.23919/JSC.2021.0018**

Greene, D., Hoffmann, A. L., & Stark, L. (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* (pp. 2122-2131). AIS. **https://aisel.aisnet.org/...**

Grosz, B. J., Grant, D. G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A., & Waldo, J. (2019). Embedded EthiCS: Integrating ethics across CS education. *Communications of the ACM*, *62*(8), 54-61. **https://doi.org/10.1145/3330794**

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99-120. **https://doi.org/ggncpj**

Hanif, A., Zhang, X., & Wood, S. (2021). A Survey on explainable artificial intelligence techniques and challenges. In *Proceedings of the 2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW)* (pp. 81-89). IEEE. **https://doi.org/10.1109/EDOCW52865.2021.00036**

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*(3), 575-599. **https://doi.org/10.2307/3178066**

Harrison, G., Hanson, J., Jacinto, C., Ramirez, J., & Ur, B. (2020). An empirical study on the perceived fairness of realistic, imperfect machine learning models. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 392-402). **https://doi.org/grttjs**

Henderson, T. (2019). Teaching data ethics: We're going to ethics the heck out of this. In *Proceedings of the 3rd Conference on Computing Education Practice* (article 4). ACM. **https://doi.org/grdcsf**

Hess, J. L., & Fore, G. (2018). A Systematic literature review of US engineering ethics interventions. *Science and Engineering Ethics*, *24*, 551-583. **https://doi.org/ggbg5c**

Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *AI and Ethics*, *1*(1), 41-47. **https://doi.org/gm9css**

Hoffmann, A. L. (2019). Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, *22*(7), 900-915. **https://doi.org/gf2m8v**

Horton, D., Liu, D., McIlraith, S. A., & Wang, N. (2023). Is more better when embedding ethics in CS courses? In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education* (Vol. 1, pp. 652-658). ACM. **https://doi.org/pt64**

Horton, D., McIlraith, S. A., Wang, N., Majedi, M., McClure, E., & Wald, B. (2022). Embedding ethics in computer science courses: Does it work? In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education* (Vol. 1, pp. 481-487). ACM. **https://doi.org/pt65**

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399. **https://doi.org/gf73q2**

Kuhn, T. S. (2012). *The structure of scientific revolutions* (I. Hacking, Ed.; 4th ed.). University of Chicago Press. (Original work published 1962.)

Latour, B. (1987). *Science in action*. Harvard University Press.

Lauer, D. (2021). You cannot have AI ethics without ethics. *AI and Ethics*, *1*(1), 21-25. **https://doi.org/ghzjtb**

Lies, J. M., Bock, T., Brandenberger, J., & Trozzolo, T. A. (2012). The effects of off-campus service learning on the moral reasoning of college students. *Journal of Moral Education*, *41*(2), 189-199. **https://doi.org/pwch**

McDonald, N., & Pan, S. (2020). Intersectional AI: A study of how information science students think about ethics and their impact. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), article 147. **https://doi.org/10.1145/3415218**

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development? In G. T. Leavens, A. Garcia, & C. S. Păsăreanu (Eds.), *ESEC/FSE 2018: Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 729-733). ACM. **https://doi.org/gm5prz**

Metcalf, J., Moss, E., & boyd, d. (2019). Owning ethics: Corporate logics, Silicon Valley, and the institutionalization of ethics. *Social Research: An International Quarterly*, *86*(2), 449-476. **https://doi.org/10.1353/sor.2019.0022**

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501-507. **https://doi.org/ggdjzn**

Morin, E. (1990). *Science avec conscience* [*Science with conscience*] (New edition). Éditions du Seuil.

Morin, E. (2005). *Introduction à la pensée complexe* [*Introduction to complexity*]. Éditions du Seuil.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141-2168. **https://doi.org/ghc8ng**

Munn, L. (2022). The uselessness of AI ethics. *AI and Ethics*, *3*(3), 869-877. **https://doi.org/gqs7nz**

Nakatumba-Nabende, J., Suuna, C., & Bainomugisha, E. (2023). AI ethics in higher education: Research experiences from practical development and deployment of AI systems. In C. C. Corrigan, S. A. Asakipaam, J. J. Kponyo, & C. Luetge (Eds.), *AI ethics in higher education: Insights from Africa and beyond* (pp. 39-55). Springer. **https://doi.org/pwck**

Naphan-Kingery, D. E., Miles, M., Brockman, A., McKane, R., Botchway, P., & McGee, E. (2019). Investigation of an equity ethic in engineering and computing doctoral students. *Journal of Engineering Education*, *108*(3), 337-354. **https://doi.org/10.1002/jee.20284**

Noddings, N. (2005). *The challenge to care in schools: An alternative approach to education* (2nd ed). Teachers College Press.

Novak, J. M., Markey, V., & Allen, M. (2007). Evaluating cognitive outcomes of service learning in higher education: A meta-analysis. *Communication Research Reports*, *24*(2), 149-157. **https://doi.org/fqrgsv**

Puig de la Bellacasa, M. (2017). *Matters of care: Speculative ethics in more than human worlds*. University of Minnesota Press.

Raji, I. D., Scheuerman, M. K., & Amironesei, R. (2021). You can't sit with us: Exclusionary pedagogy in AI ethics education. In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 515-525). ACM. **https://doi.org/gh7f5p**

Reich, R., Sahami, M., Weinstein, J. M., & Cohen, H. (2020). Teaching computer ethics: A deeply multidisciplinary approach. In *SIGCSE '20: Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 296-302). ACM. **https://doi.org/ggxqbf**

Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*, *7*(2). **https://doi.org/gg9ndb**

Rulifson, G., & Bielefeldt, A. (2017). Motivations to leave engineering: Through a lens of social responsibility. *Engineering Studies*, *9*(3), 222-248. **https://doi.org/gc3hfj**

Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., Dewar, N., & Beard, N. (2019). Integrating ethics within machine learning courses. *ACM Transactions on Computing Education*, *19*(4). **https://doi.org/10.1145/3341164**

Schiff, D. (2021). Education for AI, not AI for Education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*, *32*(3), 527-563. **https://doi.org/gs7d89**

Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2021). Explaining the principles to practices gap in AI. *IEEE Technology and Society Magazine*, *40*(2), 81-94. **https://doi.org/10.1109/MTS.2021.3056286**

Schwartz Cowan, R. (1985). *More work for mother: The ironies of household technology from the open hearth to the microwave*. Basic Books.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *FAT* '19: Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency* (pp. 59-68). ACM. **https://doi.org/gftmp3**

Shapiro, B. R., Lovegall, E., Meng, A., Borenstein, J., & Zegura, E. (2021). Using role-play to scale the integration of ethics across the computer science curriculum. In *SIGCSE '21: Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 1034-1040). ACM. **https://doi.org/pwcr**

Shapiro, B. R., Meng, A., O'Donnell, C., Lou, C., Zhao, E., Dankwa, B., & Hostetler, A. (2020). Re-shape: A method to teach data ethics for data science education. In *CHI '20: CHI Conference on Human Factors in Computing Systems*. ACM. **https://doi.org/ghrmf5**

Shen, H., Deng, W. H., Chattopadhyay, A., Wu, Z. S., Wang, X., & Zhu, H. (2021). Value cards: An educational toolkit for teaching social impacts of machine learning through deliberation. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 850-861). ACM. **https://doi.org/gh73fv**

Slack, J. D., & Wise, J. M. (2015). *Culture + technology: A primer* (2nd ed.). Peter Lang.

Stark, L., Greene, D., & Hoffmann, A. L. (2021). Critical perspectives on governance mechanisms for AI/ML systems. In J. Roberge & M. Castelle (Eds.), *The cultural life of machine learning* (pp. 257-280). Palgrave Macmillan. **https://doi.org/pwcs**

Stark, L., & Hoffmann, A. L. (2019). Data is the new what? Popular metaphors & professional ethics in emerging data culture. *Journal of Cultural Analytics*, *4*(1). **https://doi.org/10.22148/16.036**

Tuovinen, L., & Rohunen, A. (2021). Teaching AI ethics to engineering students: Reflections on syllabus design and teaching methods. In J. Koskinen, M. M. Rantanen, A.-M. Tuikka, & S. Knaapi-Junnila (Eds.), *Proceedings of the Conference on Technology Ethics 2021 – Tethics 2021* (pp. 19-33). CEUR. **https://ceur-ws.org/Vol-3069/FP_02.pdf**

Villegas-Galaviz, C., & Martin, K. (2023). Moral distance, AI, and the ethics of care. *AI & Society*, *39*(4), 1695-1706. **https://doi.org/pwcv**

Waddington, D. I., & Weeth Feinstein, N. (2016). Beyond the search for truth: Dewey's humble and humanistic vision of science education. *Educational Theory*, *66*(1-2), 111-126. **https://doi.org/10.1111/edth.12157**

Wagner, B. (2018). Ethics as an escape from regulation: From "ethics-washing" to ethics-shopping? In E. Bayamlioğlu, I. Baraliuc, L. Janssens, & M. Hildebrandt (Eds.), *Being profiled: Cogitas ergo sum – 10 years of profiling the European citizen* (pp. 84-89). Amsterdam University Press. **https://doi.org/10.2307/j.ctvhrd092.18**

Weiss, A., Vrecar, R., Zamiechowska, J., & Purgathofer, P. (2021). Using the design of adversarial chatbots as a means to expose computer science students to the importance of ethics and responsible design of AI technologies. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, & K. Inkpen (Eds.), *Human-Computer Interaction – INTERACT 2021* (Lecture Notes in Computer Science, Vol. 12934, pp. 331-339). Springer. **https://doi.org/pwcw**

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating systems: Gender, race and power in AI*. AI Now Institute. **https://ainowinstitute.org/discriminatingsystems.pdf**

Williams, R. M., Smarr, S., Prioleau, D., & Gilbert, J. E. (2022). Oh no, not another trolley! On the need for a co-liberative consciousness in CS pedagogy. *IEEE Transactions on Technology and Society*, *3*(1), 67-74. **https://doi.org/10.1109/TTS.2021.3084913**

Winner, L. (2020). *The whale and the reactor: A search for limits in an age of high technology* (2nd ed.). The University of Chicago Press.