Integrating Ethics into Computer Science Education

Multi-, Inter-, and Transdisciplinary Approaches

Trystan S. Goetze
tgoetze@fas.harvard.edu
Harvard University
Cambridge, Massachusetts, USA

ABSTRACT

While calls to integrate ethics into computer science education go back decades, recent high-profile ethical failures related to computing technology by large technology companies, governments, and academic institutions have accelerated the adoption of computer ethics education at all levels of instruction. Discussions of how to integrate ethics into existing computer science programmes often focus on the structure of the intervention—embedded modules or dedicated courses, humanists or computer scientists as ethics instructors—or on the specific content to be included—lists of case studies and essential topics to cover. While proponents of computer ethics education often emphasize the importance of closely connecting ethical and technical content in these initiatives, most do not reflect in depth on the variety of ways in which the disciplines can be combined. In this paper, I deploy a framework from cross-disciplinary studies that categorizes academic projects that work across disciplines as multidisciplinary, interdisciplinary, or transdisciplinary, depending on the degree of integration. When applied to computer ethics education, this framework is orthogonal to the structure and content of the initiative, as I illustrate using examples of dedicated ethics courses and embedded modules. It therefore highlights additional features of cross-disciplinary teaching that need to be considered when planning a computer ethics programme. I argue that computer ethics education should aim to be at least interdisciplinary—multidisciplinary initiatives are less aligned with the pedagogical aims of computer ethics—and that computer ethics educators should experiment with fully transdisciplinary education that could transform computer science as a whole for the better.

CCS CONCEPTS

• **Social and professional topics** → **Computer science education**; *Codes of ethics.*

KEYWORDS

ethics education, embedded ethics, data justice, ethics course, higher education, cross-disciplinary studies, interdisciplinary teaching and learning, responsible computing, transdisciplinary studies, interdisciplinary studies

SIGCSE 2023, March 15–18, 2023, Toronto, ON, Canada.

ACM ISBN 978-1-4503-9431-4/23/03…$15.00
https://doi.org/10.1145/3545945.3569792

ACM Reference Format:
Trystan S. Goetze. 2023. Integrating Ethics into Computer Science Education: Multi-, Inter-, and Transdisciplinary Approaches. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2023), March 15–18, 2023, Toronto, ON, Canada.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3545945.3569792

1 INTRODUCTION

It is more apparent than ever that technical education is incomplete without ethics education. Calls to integrate some kind of ethics instruction into computer science education in particular have been made for decades, but a spate of tech scandals in recent years—whether by tech companies [22], governments [12], or academic institutions [4]—has accelerated the adoption of computer ethics into existing programmes of study.

Recognizing that there is a need for ethics in computer science education is one thing; devising and implementing strategies for integrating such instruction is another. But there is a consistent thread throughout the history of computer ethics education: students should encounter ethical and technical topics in tandem. Writing in 1988, Miller contends that "the societal and technical aspects of computing are interdependent. Technical issues are best understood (and most effectively taught) in their social context, and the societal aspects of computing are best understood in the context of the underlying technical detail" [29, p. 38]. However, the disciplinary structures of modern institutions of higher education interfere with such integration. Expertise in ethics education and expertise in technical education are usually developed by distinct progammes of study, in humanistic, social-scientific, and area studies departments on the one hand, and in science, technology, engineering, and mathematics (STEM) departments on the other. As a result, computer ethics education is necessarily cross-disciplinary.

Much has been written on cross-disciplinary work in the academy [2, 15, 21, 26, 33]. In this paper, I provide a specific framework from cross-disciplinary studies to help organize the development of computer ethics curricula. The project of the paper is thus a sort of conceptual engineering, that is to say, a philosophical intervention that seeks to clarify and improve the concepts we use to structure our thinking [7, 8, 20]. Terms like "multidisciplinary," "interdisciplinary," "transdisciplinary," and so on are often used interchangeably, without reflecting on differences these concepts might capture, and without explicit reflection on *how* the different disciplines in question are to be integrated. By fixing the meaning of these terms in place, we can see more clearly the particular advantages and challenges of multiple approaches to computer ethics education, and even imagine new ways of doing so.

The paper proceeds as follows. First, I briefly outline some learning goals for any computer ethics education initiative. Second, I briefly go over some typologies and frameworks for categorizing cross-disciplinary research and education, before introducing my preferred framework. This conceptual scheme places cross-disciplinary projects on a continuum from multidisciplinary to interdisciplinary to transdisciplinary, corresponding to the degree to which the disparate disciplines are integrated. I then illustrate the first two categories with successful computer ethics interventions, and I comment on the advantages and challenges of each. Finally, I discuss what transdisciplinary computer ethics education may look like, and how such an initiative may represent a transformation of the field of computing itself. My contention throughout is that computer ethics education should aim to be at least interdisciplinary, but ideally should strive towards transdisciplinarity.

A few notes before beginning. First, this paper is not a thorough literature review of relevant cross-disciplinary studies. I will only highlight a few influential definitions before developing the framework that I prefer. Second, I will not provide an argument that ethics education *should* be included in computer science curricula. I take this as sufficiently established by prior pedagogical work and by the general state of the world. Third, I will not survey specific topics that computer ethics education might cover. The particular examples of computer ethics topics that I use are not intended to be exhaustive or complete. (Instead, see [37].) The present work is focused on the form rather than the content of computer ethics education.

2 GOALS OF COMPUTER ETHICS EDUCATION

To orient us, let's start by laying out some generic learning goals for computer ethics education. These criteria will help us to judge between different approaches to integrating ethics into computer science education.

The following are drawn from the learning outcomes listed in CS2013, the ACM and IEEE's curriculum recommendations, under the Social Issues and Professional Practice area of the body of knowledge considered essential to all computer science programmes at the secondary level and higher [37]. For brevity, I have synthesized the finer-grained items in CS2013 into more general learning outcomes, and omitted those concerned with history, law, and economics.

A graduate of a computer science programme should:

(1) Be familiar with their own ethical values, those of the computing profession, those of their society, and the variety of value systems that exist in the world.
(2) Be able to analyse and produce ethical arguments.
(3) Be able to describe the positive and negative ways in which a variety of computing technologies impact the environment, individual people, and groups of people, especially marginalized social groups.
(4) Be able to identify the values and assumptions embedded in a variety of technologies.
(5) Be familiar with the values, codes, and standards of professionalism expected of a computing practitioner.

As we will see, even in this simplified schema, several of these learning goals—namely, (3), (4), and (5)—have implications for how ethics and computing should be integrated.

3 CROSS-DISCIPLINARY TYPOLOGIES

Computer ethics education may be a cross-disciplinary project, but what exactly does that entail? It seems straightforward enough that philosophical ethics (and other normative disciplines) and computer science are to be integrated in some way. However, there are many potential ways this could be accomplished. The injection of ethics might be minimal, such as a one-off lesson on the social impacts of technology in an introductory course, and still would seem to be cross-disciplinary in some sense. Or the integration of ethics and computing could be so thoroughgoing that the programme of study could be described as a social justice degree that makes heavy use of computing skills. Or the curriculum could fall somewhere in between. We need to clarify our conceptual framework to structure our reflections on such developments.

Multiple definitions of cross-disciplinary education and research exist in the literature. Unfortunately, many of them do not take us further than the self-evident fact that these initiatives combine elements of different disciplines of study. Klein remarks on the plurality of meanings of *interdisciplinarity*, noting that the common thread is that interdisciplinary work "is a means of solving problems and answering questions that cannot be satisfactorily addressed using single methods or approaches" [26, p. 196]. Similarly, Boix Mansilla describes interdisciplinary understanding as a learning outcome for students, where the goal is for students to combine insights from multiple disciplines to produce novel insights, understandings, or knowledge [6]. These definitions are helpful in that they point us towards the idea the cross-disciplinary learning and research should go beyond the limitations of any one discipline on its own. But which aspects of the disciplines should be combined and to what degree are unclear.

An early typology, developed for an OECD report, identifies a *discipline* as "a specific body of teachable knowledge with its own background of education, training, procedures, methods and content areas" [2, p. 25]. Multi- and pluridisciplinary work juxtapose multiple disciplines, where the former puts seemingly disconnected disciplines together (such as history and mathematics) and the latter juxtaposes disciplines that appear to have some affinity (such as French and Latin). Interdisciplinary work, by contrast, requires interaction between the disciplines through their distinctive concepts, methods, results, data, and so on. Finally, transdisciplinary work provides a meta-framework of concepts for understanding different disciplines from the perspective of philosophy or sociology of knowledge. This typology already recognizes that cross-disciplinary work can be shallow or deep, but does not provide much clarity on what the shallower forms may look like, or whether there are different grades of interdisciplinary integration.

Boden provides a six-tiered framework [5]. The lower tiers put multiple disciplines under one organizational umbrella and make varying formal and informal efforts to encourage communication between researchers and teachers from different backgrounds. The highest tier, what Boden calls *integrated interdisciplinarity*, requires concepts, theories, and results from one discipline to actively contribute to the other and, ideally, vice versa. She goes so far as to characterize this tier as "the only true interdisciplinarity," labelling the lower tiers "intellectually tolerant forms of multidisciplinarity" [5, p. 20]. While useful, for my purposes Boden's typology is too-fine grained on one end of the spectrum of cross-disciplinary work, and stops short of an even higher tier of disciplinary integration—what I will call *transdisciplinarity*—which I think is especially important to consider in computer ethics education.

More recently, Klein has developed a more complex typology of cross-disciplinary projects [27]. Her account largely follows that of the OECD report [2], while synthesizing it with more recent work, including some of the above. She also discusses several trends in how academics understand the notion of transdisciplinarity; these vary from frameworks that systematically integrate knowledge into an organized whole, to synthesizing paradigms that cut across multiple discipines through common elements or themes, to critical projects that aim to dismantle the very idea of disciplines of study, to approaches that shift the focus from disciplines of study to difficult societal problems that require diverse methods and perspectives to solve. Klein declines to take a stand on which understanding of "transdisciplinary" is preferred, so while her typology provides greater precision, it does not take us much further than the other works summarized above.

Writing about AI ethics education, Raji et al. argue that we should resist several common trends in teaching social and ethical issues in computing, namely, an implicit hierarchy placing technical skills as more important than ethical skills, and the false impression that good technologists are "unicorns" who can do all the technical and ethical work by themselves [32]. Instead, Raji et al. argue that computing educators must collaborate across disciplines and communities to deliver socially responsible education, calling this approach *transversal*. But Raji et al. do not provide a clear definition of "transversal." A transversal problem, on their view, "involv[es] methods, theories and collaborators across several traditional disciplines" and yet "is distinct from an interdisciplinary problem as its solution is not found in-between given disciplines" but should emerge from stakeholder-centred critiques of the impacts of technologies [32, p. 523]. The problems they mention are important, as is the inclusion of lay publics through a stakeholder lens. But I find that the notion of "transversality" remains obscure.

The framework I propose for understanding computer ethics education comes from Rosenfield [25, 34]. She uses the terms multidisciplinary, interdisciplinary, and transdisciplinary to describe three *degrees* to which cross-disciplinary work integrates the concepts, methods, and knowledge of distinct disciplines. In that order, these categories compose a scale from the least to the most integration. I prefer this framework because it combines (or is at least compatible with) the insights of the approaches cited above, while also going beyond them to include a specific definition of transdisciplinarity that marks a higher level of disciplinary integration. In context, Rosenfield uses this framework to understand differences between several cross-disciplinary health science projects.

It is not entirely clear how Rosenfield arrived at this particular typology (she does not cite any earlier work in cross-disciplinary studies). But it is probable that she was at least indirectly influenced by the OECD report [2], given its prominence as a starting point for the usage of these terms [27]. In the following section, I provide more detail on this particular way of carving up these concepts.

4 A CROSS-DISCIPLINARY FRAMEWORK

A *multidisciplinary* collaboration involves researchers who "work in parallel or sequentially from disciplin[e]-specific base[s] to address [a] common problem" [34, p. 1351]. In these collaborations, researchers from different disciplines recognize that they have a common interest in some subject of research, and would mutually benefit from multiple inquiries from different disciplinary approaches. However, each researcher or team engages independently in the inquiry, from within their own disciplinary context, using their discipline's methods, background knowledge, and conceptual frameworks, reporting their findings to their collaborators in other disciplines only after their own work is complete. The inquiries happen either in parallel, or sequentially with one discipline's experts handing off their results to the next.

An *interdisciplinary* research project involves researchers who "work jointly but still from disciplin[e]-specific bas[es] to address [a] common problem" [34, p. 1351]. In contrast with multidisciplinary research, interdisciplinary collaborations require all disciplinary experts to work on the same project, instead of pursuing discrete inquiries. In doing so, however, each disciplinary expert or team still employs their own discipline's methods and conceptual frameworks. As in multidisciplinary research, experts from different disciplines might work sequentially, one discipline's findings being reported to the next. But instead of being essentially separate inquiries on the same subject that are brought together only at critical milestones, in an interdisciplinary collaboration, every stage of work informs the next, regardless of its disciplinary home.

A *transdisciplinary* research project involves researchers who "work jointly using [a] shared conceptual framework drawing together disciplin[e]-specific theories, concepts and approaches to address [a] common problem" [34, p. 1351]. In a transdisciplinary project, the collaboration extends beyond contributing to the same project using discrete disciplinary methods: the nature of the project requires integrating at the level of the conceptual frameworks and methods used in order to address a complex subject. In the process, a distinctive approach may emerge that could become a new discipline itself, such as the rapprochement of philosophy, psychology, neuroscience, computer science, linguistics, and other fields that produced the discipline of cognitive science [36].

To illustrate, suppose a group of philosophers and data scientists are interested in issues of data bias. In a multidisciplinary collaboration, the data scientists would work on technical issues and the philosophers on ethical issues largely independently, though their results may productively inform one another's work. In an interdisciplinary collaboration, the philosophers and data scientists would work on the same project, contributing insights and frameworks from both disciplines to inform the direction of research. In a transdisciplinary collaboration, those frameworks would instead be combined, perhaps producing a different paradigm for doing data science that makes ethics and justice equally important to success criteria as technical criteria such as statistical significance.

In the following sections, I apply this framework to computer ethics education.

5 MULTIDISCIPLINARY COMPUTER ETHICS

A multidisciplinary computer ethics initiative integrates ethical education in parallel or in sequence with technical instruction, but keeps these disciplines separate within the degree programme as a whole. Ethical concepts and skills are taught separately from technical concepts and skills, and they are brought together only at key milestones in the course of study. Ethics is treated as something that can inform and improve technical computing work, but not as essential to the work of computing per se.

One form of multidisciplinary ethics intervention is to include some aspect of ethical thinking into existing assignments in a technical course, without providing specific instruction on ethical issues from a humanistic discipline. For example, a project-based course might require students to write a section reflecting on ethical implications of their work in a report accompanying their final project, perhaps on the model of ethics statements that are now required of many computer science conference submissions. To provide some structure, this assignment might take after a disclosure tool, such as model cards [30] or dataset nutrition labels [14].

Another multidisciplinary approach is a dedicated computer ethics course. Ethics instruction takes place in parallel and in sequence with technical computer science instruction. Students are left to integrate ethics into their professional practice once they transition to the workforce or further education. During their course of study, they might not encounter ethical instruction or assessment again, unless they take other ethics or professionalism courses. I have taught such a course: Social, Ethical, and Professional Issues in Computer Science, a required course for computer science majors at Dalhousie University [17]. The topics we discussed included the ACM Code of Ethics and Professional Conduct, philosophical ethics, privacy, intellectual property, digital divides, computing and the economy, bias in computer systems, and AI ethics.

A multidisciplinary ethics intervention has some advantages, chiefly in that it requires only minimal modification to existing technical curricula. A dedicated ethics course could take the place of a breadth requirement elective, while still serving some of the same pedagogical purposes. Such a course can be taught by a single academic hire, possibly even part-time for smaller computer science programmes. Adding ethical reflection to assignments in technical courses is an even lighter lift.

There are, however, significant disadvantages to a multidisciplinary ethics intervention. These approaches fail to take on Miller's insight that technical and ethical issues are best taught in connection with one another, rather than separately. This artificially isolates subjects that, in the professional world, always co-occur: computing always takes place in a social context, and ethical issues arise in connection with specific aspects of technical computing work. As a result, it is more difficult to achieve learning goals (3) and (4). It is possible to teach students how to describe the ethical effects of computing technologies and the values embedded in technologies as abstract skills, but this will be more impactful if these skills are shown to be directly relevant to the students' own work.

Furthermore, multidisciplinary computer ethics education risks presenting ethics as supplemental to the "real" work of computer science, namely, the technical skills [19, 32]. With the bulk of the curriculum dedicated to technical instruction independent of ethics instruction, and ethics interventions limited to one-off courses or assessments, students are given the impression that ethics is an afterthought, burden, or box to tick. This impression is antithetical to the cultivation of professional responsibility among computing practitioners captured by learning goal (5).

For these reasons, multidisciplinary computer ethics instruction is less than ideal—although I speak from experience when I claim that it can still be impactful, and is worth doing if no alternative is possible. For the pedagogical goals of a computer ethics to be fully met, though, the disciplines need to be more tightly integrated.

6 INTERDISCIPLINARY COMPUTER ETHICS

A closer integration of ethics and computing education would produce an interdisciplinary programme, where both disciplines are taught concurrently in service of the same shared goals. Each discipline remains distinct in the concepts, background, and skills that they bring to bear, but both sets are used by instructors and students to engage with the same material. In interdisciplinary computer ethics education, ethics is a skillset that is presented as a requirement for doing quality work in the computing professions, despite being different in kind from technical skills.

Interdisciplinary computer ethics is perhaps easier to achieve in the context of ethics modules which are taught in the midst of ongoing technical courses. Indeed, it is explicitly our goal in the Embedded EthiCS™ programme at Harvard University, and in other initiatives that have used Embedded EthiCS as a model. As Grosz et al. explain, "For students to succeed at learning not only how to build innovative computing systems, but also how to determine whether they should build those systems or how ethical considerations should constrain their design, it is imperative that [philosophy and computing] work together" [19, p. 57]. Much of the work in the Teaching Lab, which develops and workshops ethics modules every semester, is to ensure that the ethical content of the lesson and the technical content of the course are closely connected, such that students can understand the relevance of the philosophical concepts, theories, and arguments they study in the module to their broader work in the course and in their professional careers.

For example, in a recent Embedded EthiCS module in an upper-division undergraduate course on economics and computation, the teaching fellow designed a lesson demonstrating how social and ethical issues are connected with mechanism design in a particularly vivid way [28]. The course's technical content is concerned, among other things, with formalizations of decision-making processes in the economy, such as game theory, and how those mathematical approaches can be operationalized in computational systems. The module concentrated on how mechanisms can be "strategy-proofed," such that it is no longer possible to deploy certain rational choice strategies that give some agents an advantage over others. The ethical implications of strategy-proof mechanisms were illustrated through an interactive exercise modelled on a real-world case from the Boston schools system [1]. Students took on the role of parents indicating their preferred choices of schools for their children. Students could instantly see the differences in results between the original and strategy-proofed systems, and reflected on how the two systems might impact different demographic groups.

In this example, the ethical and technical content of the module are distinct, but mutually reinforcing in a way not quite attainable with a multidisciplinary approach. The technical content—game theory, rational choice strategies, mechanism design—flows directly from the main content of the course. The ethical issue arises in connection with the presence or absence of a specific technical feature—strategy-proofing—but is explained using philosophical concepts—fairness, equality of opportunity, social justice. The potential solutions to the ethical issue are rooted in both the technical and ethical aspects of the case, and the case invites further reflection on the social conditions which produce the inequalities at issue—why, for instance, are there are better and worse schools to choose from, anyway? Because the two disciplines are more closely integrated, the relevance of social and ethical considerations to technical decisions is more clear, and the interplay between ethical reasoning and technical design processes becomes obvious.

The disadvantages of an interdisciplinary approach to computer ethics are mainly institutional and structural. Few scholars are specialists in both computing and philosophy (or some other normative discipline), so the process of developing interdisciplinary modules or courses usually must be collaborative. This can require additional time on the part of faculty and graduate students, and may require additional investment on the part of institutions. Some institutions may lack the resources or academic infrastructure; for example, some smaller colleges may not have philosophical expertise available. Such institutions should receive support from more well-resourced institutions, perhaps in the form of upskilling sessions for computer science instructors interested in teaching ethics modules. In either case, this presents another challenge for interdisciplinary work: making all parties familiar with the jargon and background knowledge of the different disciplines.

The main advantages of an interdisciplinary computer ethics approach are pedagogical. As we saw, in a multidisciplinary initiative, the separation of the disciplines is in tension with learning goals (3), (4), and (5). To meet these goals, an interdisciplinary initiative enables students to engage in ethical evaluation of the same technologies that they are studying in their technical courses. The assumptions of the developers of these technologies and the values embedded in those technologies become more salient when the technical features of those technologies are already top-of-mind. And, the inculcation of professional responsibility is facilitated by engaging with these questions in a similar context to that in which ethical issues might arise in professional practice, namely, while in the midst of learning about and deploying a technical system.

Teaching communities that are serious about including ethics in a computing curriculum should therefore strive to achieve at least an interdisciplinary approach, despite its potential challenges. The learning goals cannot be met as effectively by a multidisciplinary intervention—though multidisciplinary computer ethics is a reasonable fallback where resources are unavailable.

7 TRANSDISCIPLINARY COMPUTER ETHICS

The prefix "trans-" in *transdisciplinary* suggests that such a project *transcends* the conventional divide between the disciplines involved, going over and beyond their limitations and perhaps creating a new field of study better equipped to tackle some range of problems. For example, one way to interpret the emergence of cognitive science and of science and technology studies as distinctive fields is that the various disciplines studying the mind and the social aspects of technoscience respectively came together to create new concepts, methods, and theories that could not be produced within their progenitor disciplines [24, 36]. Similarly, we might wonder what may emerge if ethics and computing were to become so closely integrated as to produce a new discipline of research and education.

Transdisciplinary work requires that practitioners in multiple disciplines use their distinct expertises to craft a new conceptual and methodological framework to tackle a shared inquiry. Fully detailing such a framework for computer ethics is beyond the scope of the present paper, but we can get a sense of what transdisciplinary computer ethics might look like by examining the emerging *data justice* movement. As D'Ignazio and Klein describe it, one facet of data justice is to move away from post hoc ethical fixes—such as correcting data bias only *after* downstream problems emerge—and instead to consider social justice at every step of the design, development, and deployment of a system [11].

This approach to computer ethics is one facet of a more radical conceptual shift that D'Ignazio and Klein recommend, away from conceiving of ethical failures in technology as merely isolated incidents caused by immoral individuals or broken technologies, and towards a more systemic understanding of the causes of harm and injustice associated with various sociotechnical systems. We might capture this insight as aligning technical and ethical criteria for a successful project: the system as a whole must function well in the actual context of use, both in terms of operating without technical fault *and* in terms of avoiding harm and promoting good.

This last point is one I want to emphasize: *a transdisciplinary approach to computer ethics would make ethical success at least as important as technical success in computing generally.* On this approach, ethical reflection is no longer a supplemental consideration, as multidisciplinary interventions position it, nor is it an allied but still alien skillset, as it is positioned by interdisciplinary initiatives. Rather, on a transdisciplinary vision, ethical reflection is a core skill for computing, as important as other specialized areas of computing practice. A technological innovation would not be considered *technically* sound unless it is also *ethically* responsible to deploy. The conceptual frameworks are not just running in parallel; they have merged.

How could transdisciplinary computer ethics be achieved in a computer science education context? Here is a speculative outline of some elements of a transdisciplinary Introduction to Computer Science course. Social and ethical issues would be raised in conjunction with technical topics at each step of the way. Introduction of the basic functions of a computer and computing hardware are presented in conjunction with a discussion of sustainability of natural resources and labour justice in manufacturing [10]. Discussion of different programming languages raises issues about the usability of different languages, with a focus on accessibility [38]. Discussion of efficiency as a technical criterion for success in programming includes ethical concerns about the usage of limited computing resources and the massive energy requirements of some computing applications, such as large language model training [35]. Assignments to edit and correct code require not just consideration of syntax errors and failures to fulfill the requirements of a project brief, but also the social impacts of poor design choices [39].

More advanced computing courses might push the integration of ethics and computing further. A transdisciplinary data science course, for example, might start from the premise that data science can and should be used to further the good of all, and implement a project-based pedagogy where students do just that. Everything they learn in such a course—technical data science skills, yes, but also theories of social justice, debiasing methods, and/or community engagement practices—might lead to a final project intended to address some inequity in their local community, perhaps using public datasets to create visualizations of important and under-discussed trends, such as differential health outcomes based on local factors.

Even highly technical computing courses can incorporate ethical concerns. A course on compilers, for instance, might again consider the energy requirements of different computing languages [31]. Or, because compilation touches source code and machine code, and involves distinct pieces of software that may be released under different licences, various stages of a compilers course may raise discussion of the ethics of intellectual property and free and open-source software [16]. It may be more difficult to incorporate discussion of ethical issues throughout theory-heavy courses in computing, but interdisciplinary modules can still provide a link back to the transdisciplinary programme's overall merging of the ethical and the technical. For example, a module in a theory-heavy course might consider whether theorists bear any responsibility for the downstream applications of their ideas [18].

With sufficient commitment from faculty in computing, ethics, and other disciplines, a transdisciplinary computing endeavour might scale up into a distinctive minor or major concentration in an emergent field, say, data justice or responsible computing. However, this raises the main disadvantage of a transdisciplinary approach to teaching and research: even more than interdisciplinary projects, transdisciplinary work requires significant investment of time, resources, and personnel to be successful. What is more, academic incentives may act against transdisciplinary work, requiring administrators to cooperate in redefining hiring, retention, and tenure criteria. It may also be a difficult sell to more traditionally minded faculty who believe that a computer science degree should focus on getting students to master a wide range of technical skills to a high degree of ability. These are common challenges across transdisciplinary endeavours, however. Considering the successes of transdisciplinary projects in other fields may be instructive to those seeking an ethical transformation of computer science.

The potential advantages of a transdisciplinary computer ethics education are significant. A transdisciplinary computer science and computer ethics education—whether we call it data justice, responsible computing, or something else—stands perhaps the best chance of creating technologists who are well-rounded people in the sense striven for by liberal arts programmes, while also providing graduates with a robust set of computing skills to enable them to make positive change with their innovations. In a way, what a transdisciplinary computer ethics education proposes is a twenty-first century version of the successful person that Aristotle tells us is possible only by cultivating the various virtues of character and intellect [3, 23]. A rough-and-ready summary of Aristotle's ethics is that to live a successful life, one must cultivate virtues of character (e.g. generosity, courage, wit), virtues of intellect (e.g. knowledge, skill, wisdom), a commitment to justice and civic duty, and good relations with friends and family. Aristotle thinks this is best achieved in a life of contemplation—that is to say, of studying philosophy among like-minded fellows. But a transdisciplinary responsible computing education—and the transformed computing professions that graduates thereof may help to produce—could well be a modern alternative [cf. 13, pp. 16–17].

This raises the worry from Raji et al., mentioned earlier, that computer ethics education can give the false impression that technologists should be able to "do it all," what they describe as "the engineer's natural inclination towards seeing themselves as a solitary saviour, to the detriment of the quality of the solution and in spite of the need for other disciplinary perspectives" [32, p. 515]. They argue against this view of the ideal computing professional as an "ethical unicorn" who can do all the technical and ethical work alike, contending that the sorts of ethical issues that computing professionals now face are more often structural problems that require structural solutions.

However, we may note that computing is already a large and complex assemblage of multiple specializations; to make a technology well requires a team of experts in several different subfields. So too with responsible computing and data justice. A transdisciplinary computer ethics education programme would be successful were it to reliably produce experts in computer ethics and experts in other subfields of computing, each of whom have enough complementary technical or ethical knowledge to be interactional experts [9] with the others. A certain baseline level of ethical skill would be expected of all, but there would remain an important role for computer ethics specialists, whose backgrounds might be subdivided further into those who focus on philosophy, social science, area studies, or stakeholder engagement.

8 CONCLUSION

In this paper, I outlined a framework from cross-disciplinary studies and used it to analyse three different approaches to teaching computer ethics on the basis of how closely the disciplines of ethics and computing are integrated. I argued that any computer ethics curriculum should aim to be at least interdisciplinary, or else it will fail to meet some of the central learning goals of computer ethics. I also sketched possibilities for transdisciplinary computer ethics, which has the potential to launch a new and distinct form of the computing profession or a specialization within it.

ACKNOWLEDGMENTS

Thanks to Alison Simmons, Barbara Grosz, Jeff Behrends, Salil Vadhan, James Mickens, Stephen Chong, Jenna Donohue, William Cochran, Jill Susarrey, and everyone else in Embedded EthiCS @ Harvard for the collegiality and support that enabled this paper. Thanks also to Casey Fiesler, Marianna Ganapini, Kristina Striegnitz, Letitia Meynell, Travis LaCroix, three anonymous reviewers for SIGCSE, and audiences at Union College and Dalhousie University, for conversation about these ideas. This paper was written on the traditional and ancestral territory of the Massachusett people.

REFERENCES

[1] Atila Abdulkadiroglu, Parag Pathak, Alvin E. Roth, and Tayfun Sonmez. 2006. *Changing the Boston School Choice Mechanism.* Technical Report. NBER Working Paper Series. http://www.nber.org/papers/w11965

[2] Léo Apostel, Guy Berger, Asa Briggs, and Guy Michaud (Eds.). 1972. *Interdisciplinarity: Problems of Teaching and Research in Universities.* Organization for Economic Cooperation and Development, Paris, France.

[3] Aristotle and Terence Irwin (Ed./Trans.). 1999. *Nicomachean Ethics* (2nd ed.). Hackett, Indianapolis, IN.

[4] Abeba Birhane and Vinay Uday Prabhu. 2021. Large image datasets: A pyrrhic win for computer vision?. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Piscataway, NJ, 1536–1546. https://doi.org/10.1109/WACV48630.2021.00158

[5] Margaret A. Boden. 1999. What is Interdisciplinarity?. In *Interdisciplinarity and the Organisation of Knowledge in Europe*, Richard Cunningham (Ed.). Office for Official Publications of the European Communities, Luxembourg, 13–24. https://op.europa.eu/en/publication-detail/-/publication/40ce830d-2e19-4af0-b521-863857b8b06c

[6] Veronica Boix Mansilla. 2005. Assessing Student Work at Disciplinary Crossroads. *Change: The Magazine of Higher Learning* 37 (2005), 14–21. Issue 1. https://doi.org/10.3200/CHNG.37.1.14-21

[7] Herman Cappelen. 2018. *Fixing Language: An Essay on Conceptual Engineering.* Oxford University Press, Oxford, UK.

[8] Herman Cappelen, David Plunkett, and Alexis Burgess (Eds.). 2019. *Conceptual Engineering and Conceptual Ethics.* Oxford University Press, New York, NY.

[9] Harry Collins and Robert Evans. 2007. *Rethinking Expertise.* The University of Chicago Press, Chicago, IL.

[10] Kate Crawford. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale University Press, New Haven, CT.

[11] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism.* The MIT Press, Cambridge, MA. https://data-feminism.mitpress.mit.edu/

[12] Jessica Elgot and Richard Adams. 2020. Ofqual exam results algorithm was unlawful, says Labour. https://www.theguardian.com/education/2020/aug/19/ofqual-exam-results-algorithm-was-unlawful-says-labour

[13] M. David Ermann and Michele S. Shauf (Eds.). 2003. *Computers, Ethics, and Society* (3rd ed.). Oxford University Press, New York, NY, and Oxford, UK.

[14] Kasia Chmielinski et al. 2021. Data Nutrition Project. https://datanutrition.org/

[15] Robert Frodeman (Ed.). 2017. *The Oxford Handbook of Interdisciplinarity* (2nd ed.). Oxford University Press, Oxford, UK. https://doi.org/10.1093/oxfordhb/9780198733522.001.0001

[16] Trystan S. Goetze. 2021. Embedded EthiCS Module: Compilers (CS 153) - Fall 2021. https://embeddedethics.seas.harvard.edu/classes/compilers-cs-153-fall-2021

[17] Trystan S. Goetze. 2021. Syllabus Showcase: Social, Ethical, and Professional Issues in Computer Science. https://blog.apaonline.org/2021/08/18/syllabus-showcase-social-ethical-and-professional-issues-in-computer-science-trystan-goetze/

[18] Trystan S. Goetze. 2022. Mind the Gap: Autonomous Systems, the Responsibility Gap, and Moral Entanglement. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea, and Online) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 390–400. https://doi.org/10.1145/3531146.3533106

[19] Barbara J. Grosz, David Gray Grant, Kate Vredenburgh, Jeff Behrends, Lily Hu, Alison Simmons, and Jim Waldo. 2019. Embedded EthiCS: Integrating Ethics across CS Education. *Commun. ACM* 62, 8 (July 2019), 54–61. https://doi.org/10.1145/3330794

[20] Sally Haslanger. 2012. *Resisting Reality: Social Construction and Social Critique.* Oxford University Press, Oxford, UK.

[21] Carolyn Haynes (Ed.). 2002. *Innovations in interdisciplinary teaching.* American Council on Education and The Oryx Press, Westport, CT.

[22] Melissa Heikkilä. 2022. The viral AI avatar app Lensa undressed me—without my consent. https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/

[23] D. S. Hutchinson. 1995. Ethics. In *The Cambridge Companion to Aristotle*, Jonathan Barnes (Ed.). Cambridge University Press, Cambridge, UK, 195–232.

[24] Sheila Jasanoff. 2017. A Field of Its Own: The Emergence of Science and Technology Studies. In *The Oxford Handbook of Interdisciplinarity* (2nd ed.), Robert Frodeman (Ed.). Oxford University Press, Oxford, UK, 173–187. https://doi.org/10.1093/oxfordhb/9780198733522.013.15,

[25] Frank Kessel and Patricia L. Rosenfield. 2008. Toward Transdisciplinary Research: Historical and Contemporary Perspectives. *American Journal of Preventative Medicine* 35, 2S (2008), S225–S234. https://doi.org/10.1016/j.amepre.2008.05.005

[26] Julie Thompson Klein. 1990. *Interdisciplinarity: History, Theory, and Practice.* Wayne State University Press, Detroit, MI.

[27] Julie Thompson Klein. 2017. Typologies of Interdisciplinarity: The Boundary Work of Definition. In *The Oxford Handbook of Interdisciplinarity* (2nd ed.), Robert Frodeman (Ed.). Oxford University Press, Oxford, UK, 21–34. https://doi.org/10.1093/oxfordhb/9780198733522.013.3

[28] Ellie Lasater-Guttmann. 2021. Embedded EthiCS Module: Economics and Computation (CS 136) - Fall 2021. https://embeddedethics.seas.harvard.edu/classes/economics-and-computation-cs-136-fall-2021

[29] Keith Miller. 1988. Integrating Computer Ethics into the Computer Science Curriculum. *Computer Science Education* 1, 1 (1988), 37–52. https://doi.org/10.1080/0899340880010104

[30] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3287560.3287596

[31] Rui Pereira, Marco Couto, Francisco Ribeiro, Rui Rua, Jácome Cunha, João Paulo Fernandes, and João Saraiva. 2017. Energy Efficiency across Programming Languages: How Do Energy, Time, and Memory Relate?. In *Proceedings of the 10th ACM SIGPLAN International Conference on Software Language Engineering* (Vancouver, BC, Canada) *(SLE 2017)*. Association for Computing Machinery, New York, NY, USA, 256–267. https://doi.org/10.1145/3136014.3136031

[32] Inioluwa Deborah Raji, Morgan Klaus Scheuerman, and Razvan Amironesei. 2021. "You Can't Sit With Us": Exclusionary Pedagogy in AI Ethics Education. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT'21)* (3–10). ACM, Virtual Event, Canada, 515–525. https://doi.org/10.1145/3442188.3445914

[33] Allen F. Repko and Rick Szostak. 2021. *Interdisciplinary Research: Process and Theory* (4th ed.). SAGE, Thousand Oaks, CA.

[34] Patricia L. Rosenfield. 1992. The Potential of Transdisciplinary Research for Sustaining and Extending Linkages Between the Health and Social Sciences. *Social Science & Medicine* 35, 11 (1992), 1343–1357. https://doi.org/10.1016/0277-9536(92)90038-r

[35] Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, Stroudsburg, PA, 3645–3650. https://doi.org/10.48550/arXiv.1906.02243

[36] Paul Thagard. 2017. Cognitive Science. In *The Oxford Handbook of Interdisciplinarity* (2nd ed.), Robert Frodeman (Ed.). Oxford University Press, Oxford, UK, 188–200. https://doi.org/10.1093/oxfordhb/9780198733522.013.16

[37] The Joint Task Force on Computing Curricula of the Association for Computing Machinery (ACM) and the IEEE Computer Society. 2013. *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science.* Technical Report. ACM and IEEE. http://dx.doi.org/10.1145/2534860

[38] Eliza Wells. 2022. Embedded EthiCS Module: Programming Languages (CS152) - 2022 Spring. https://embeddedethics.seas.harvard.edu/cs-152-2022-spring

[39] Nick Zufelt. 2019. Salutations Generator Assignment; Or, Refactoring for Inclusivity. https://github.com/nzufelt/salutations_generator_assignment