

Submission to CACM Research Highlights: How to Teach Computer Ethics with Science Fiction

Emanuelle Burton
University of Illinois at Chicago
Chicago, IL, USA
enburton@uic.edu

Judy Goldsmith
University of Kentucky
Lexington, Kentucky, USA
goldsmit@cs.uky.edu

Nicholas Mattei
IBM Research
Yorktown Heights, NY, USA
n.mattei@ibm.com

1. INTRODUCTION

Computer science faculty have a responsibility to teach students to recognize both the larger ethical issues and particular responsibilities that are part and parcel of their work as technologists. This is, however, a kind of teaching for which most of us have not been trained, and one which faculty and students approach with some trepidation. In this article we explore the use of science fiction as an effective tool to enable those teaching AI to engage students and practitioners about the scope and implications of current and future work in computer science.

Computer science, as a field, has already recognized that some ethics education is essential; ABET, one of the largest accreditors of engineering and technology programs, requires instruction on professional ethics. Indeed, some in CS have gone so far as to require students in undergraduate courses to perform ethics consultations for local industry [33]. However, educating students to engage with ethical challenges is often left to the cross-disciplinary portions of university curricula (especially in the US [20]). We, among others, argue that spending time focused on how these issues apply to both our own research and students' future work is important and necessary *within CS* [43, 50].

In fields with a strong practical component and established body of knowledge, e.g., medicine, engineering, and the undergraduate levels of many sciences, there is a temptation to teach by transmitting facts, rather than encourage discussion and dissent [18]. This approach, which many undergraduates have seen before entering the university, can condition students into interpreting what they learn in terms of an authority-based view of "truth," which in turn leaves them unequipped to reason about situations which involve no single correct answer, or to think cogently about ethical trade-offs [32, 48]. In computer science we want to teach our students to move past this authority-based view and find the best, most efficient solution to technical problems; we argue that the same skills must be developed to engage with ethical challenges that arise from the substance of their work as well.

We have spent several years developing a creative approach to teaching computer ethics, a course called "Science Fiction and Computer Ethics" [39, 15, 16, 14, 28].

The course has been taught four times at the University of Kentucky and once at the University of Illinois at Chicago (thrice by Goldsmith, twice by Burton). The course has been well received and will be taught in Spring 2018 at both universities.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2008 ACM 0001-0782/08/0X00 ...\$5.00.

This course has been successful with students as evidenced by increasing and full enrollments; high teaching evaluation numbers; positive anonymous comments from students; nominations and awards for good teaching; and invitations to speak about the course at conference panels and talks including the Symposium on Educational Advances in Artificial Intelligence (EAAI). A related undergraduate computer science course that uses science fiction to teach students about research has also been qualitatively and quantitatively successful [29, 30]. Many courses with ethical considerations taught through fiction have been offered in the past. Examples include ones at Humboldt University at Berlin¹, a version focused on legal issues at Stanford², among others [4, 5, 6]. Courses in other fields use literature (including science fiction) in non-majors courses as both a "hook" and a platform for exploring core issues [6, 21]. Scholars in other humanistic disciplines such as history and philosophy have also argued that literature is an invaluable teaching tool for ethics, among other topics (e.g., [26, 27, 49]). The common observation of these courses is that the fiction-based approach makes it much easier to push beyond a review of best practices and achieve a more in-depth education in ethical reasoning; "[...] fiction often removes the intellectual and emotional resistance some students might at first feel towards the subject of ethics [47]."

2. ETHICS AND VALUES IN COMPUTER SCIENCE

Among researchers in the computing professions, as in all professions, there are multiple and often conflicting sets of values, as well as different ways to approach the task of living up to one's values. It is important to be clear that the purpose of teaching ethics is not to unify the field around a particular value system, but rather to encourage reflection and precision of thought among all computer professionals. This will hopefully lead to an openness and exchange of ideas about both core values and best practices.

The very idea of a universally-applicable ethical doctrine has serious problems. As anthropologist Melville Herskovits wrote in protest of the UN's Universal Declaration of Human Rights, the declaration — although intended "to be applicable to all human beings... [is] conceived only in terms of the values prevalent in countries of Western Europe and America [56]." In other words, any attempt to codify a universal definition of the "right" way to be human cannot, by definition, take account of the particular social and ethical context of individual cultures. Therefore, cultures that have historically been most oppressed would be the most likely to be

¹<http://waste.informatik.hu-berlin.de/Lehre/ws0910/dystopien/>

²<http://web.stanford.edu/class/cs122/>

ignored or de-legitimized by any “universal” declaration.

Although the precise status and possibilities of human rights discourse is still debated, scholars in both ethics and anthropology agree that there is no way to formulate universal precepts of this kind that do not, on some level, *reinforce* the very kinds of social inequality they are designed to combat. The idea that a single code of laws or duties would solve all problems, and that our responsibility as teachers is to transmit those laws to students, is appealing but ultimately false. As Callahan writes,

No teacher of ethics can assume that he or she has such a solid grasp on the nature of morality as to pretend to know what finally counts as good moral conduct. No society can assume that it has any better grasp of what so counts as to empower teachers to propagate it in colleges and universities. Perhaps most importantly, the premise of higher education is that students are at an age where they have to begin coming to their own conclusions and shaping their own view of the world. It is the time and place to teach them intellectual independence, and to instill in them a spirit of critical inquiry [17].

Rather than ascertaining and transmitting such a law to students, the responsibility of an ethics instructor is to train the students to engage in understanding and reasoning. Thus the students are prepared to navigate situations that offer no clean solutions, and engage other CS practitioners in discussion about what and how to choose. In Callahan’s words, to “help students develop a means and a process for achieving their own moral judgments” [17] when confronted with challenging situations.

It is essential that open ethical debates between well-informed practitioners take place. Computer science does not take place in a vacuum; to an ever-increasing degree, the IT systems and platforms from search engines to smart phones that are built by computer scientists are creating and redefining the social, political, and individual contexts in which human beings understand themselves [31]. Whatever principles and norms are adopted by computer scientists, and reinforced through the design and deployment of their systems, will have profound ethical and societal implications.

Teachers and leaders in the field have a responsibility to drive the discussion about the impacts of their own work and the work of their students. Indeed, Boyer argues that academics have a responsibility to engage students and the public with their research [11, 10]. We have started to see this through a number of initiatives in the CS community including: the IJCAI 2015 letter on autonomous weapons research³ and the 2017 follow-on letter signed by CEOs of tech companies around the world;⁴ the ACM statement on Algorithmic Accountability⁵; the development of the IEEE standard for algorithmic bias considerations⁶; and new conferences and research groups focused on Fairness, Accountability and Transparency.⁷ These debates are important for shaping the direction of the field, even though they rarely result in consensus. The utility of such debates is not that they result in standardized practices,

³http://futureoflife.org/AI/open_letter_autonomous_weapons

⁴<https://futureoflife.org/autonomous-weapons-open-letter-2017>

⁵https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

⁶<http://sites.ieee.org/sagroups-7003/>

⁷<http://www.fatml.org/>

but rather that individual practitioners become more thoughtful and better-informed about their work and its long-term effects.

As in other areas of thought, this diversity of viewpoints is a strength when it can be harnessed toward a productive exchange of ideas and perspectives. An example of a productive exchange is the ongoing debate within the AI research community about the appropriate value systems on which to build AI systems. The goal of teaching ethics is to foster such debates, and equip practitioners to participate in them as productively as possible. It does so, not by imposing a value system on students, but rather by informing them about the range of ethical descriptive and evaluative tools available to them, and at the same time making them aware of the social ramifications of their work: that research, development, and implementation can be carried out in a variety of ways and for a variety of ends. CS educators can, and should, dedicate significant time to ethics education, which will enable students to make informed, thoughtful, and ethical choices about technology and its applications.

2.1 What is Ethics?

Ethics can be understood simply as the task of answering the question “what should I do?” — which is never a simple question. Ethics is comprised of both thought and practice: an organized and intentional reflection on morality and the effort to live in ways that are good, just, and/or right. Although many people use the words *ethics* and *morality* interchangeably, many of ethicists understand them to be different. One common way of drawing the distinction—and one which we find helpful—is to define “morality” as a set of values or a worldview, and “ethics” as the practice of reflecting upon those values, their foundations, and their applications [8].

There are many different, often conflicting, ways of understanding how to be moral. Sometimes, the clashes are between people who share the same fundamental premises and method of inquiry into how to be moral, but disagree about conclusions. Other times, the clashes take place between people whose basic ideas of how to answer the question of how to be moral are in conflict with one another. Most approaches to morality can be understood in terms of the three major traditions of ethical thought: deontological ethics, virtue ethics, and utilitarianism. Each of these traditions grows out of different core questions and different ways of seeing the world.

Ethics is typically understood to be *normative*: that is, aimed at establishing norms of thought, values, or conduct. This assumption is especially prevalent in professional ethics courses, which are typically used as a means to steer students’ future behavior toward a set of professionally agreed-upon values, e.g., professionalism or honesty [37, 19]. But ethics is also a tool for *description*, furnishing the decision maker with a critical framework that enables her to understand what is happening in a given situation, and what is at stake in any action she might take. The boundary between normative and descriptive functions is sometimes fuzzy: it is often the case that different details of a situation will appear salient or relevant, depending on which ethical approach one adopts. This malleability of relevant details can make ethics itself seem murky or imprecise. However, teaching our students to appreciate this difference, to understand the modes of reasoning that they or others might employ in making an ethical decision, and to move between these reasoning structures themselves, is the goal of a good ethics course.

Educating students in the descriptive functions of ethics is as important as communicating to them the professional norms of computer science. CS is a field in which everyday practice and problem-solving takes place in a context that could barely be imagined the decade before. Educators cannot predict the ethical quandaries their students will face. But with an education in ethical description,

those students will be equipped to engage in subtle and substantive ethical reasoning when new and challenging problems confront them.

2.2 Practical Challenges of Teaching Ethics

Ethics education is a notably challenging task for two reasons. The first is that, in the absence of any ideal universal ethics program, students must be taught how to approach problems *as distinct from being led to particular pre-ordained conclusions* that might narrow their vision and exclude important elements of a given problem. The second challenge is how to achieve the goal outlined by the first while overcoming the biases that students bring to the classroom.

2.2.1 Teaching How, Not What, to Think

It is important to consider what it means for us to say that we want to inform our students *how* to think instead of *what* to think. It is tempting to assume that we can formulate a set of rules in natural language, refine them until we agree on them, and then proceed as if these rules can be applied without further reflection. However, the real world is messy, and rules that may seem reliable under one set of conditions can falter under others. Furthermore, language is not always identical to the world that it describes. Different people will describe the same experience in different ways, or understand the same phrase to refer to different phenomena. At a minimum, such universal rules would require everyone who relied on them to engage in ongoing reflection about their own understanding and application of these rules to the world.

Both the appeal of this rule-based approach and its limits can be seen with respect to the question of programming robots against concrete actions, e.g., programming law enforcement robots never to shoot humans. While this operating principle seems at first like a straightforward way to ensure the preservation of human life, it is not hard to imagine scenarios in which shooting a person, perhaps even lethally, will save the lives of others. But how should a robot calculate the risks and values at stake in such a scenario? What sorts of input should it use when ascertaining if it should shoot a human, and what sorts of input should it ignore? And what are the social costs or benefits of using robots that will shoot a human under certain circumstances? Another example is the extended recent discussion about the classic trolley problem in light of the rapid advance of self driving cars [9]

Ethics education often requires a different kind of education than understanding and applying an established body of knowledge. In computer science, knowledge usually leads to action: if you choose to create or program a system to solve a problem, and know how to do so, there is very little reason not to solve the problem in the most direct and efficient way possible. Ethical understanding, however, requires an additional layer of commitment. One must overcome both the temptations to adopt an easier or more self-serving course and the distractions that might prevent someone from recognizing an ethical problem in the first place. It is not hard to imagine a student who can get 100% on a class exam, correctly identifying terms and offering cogent and sensible solutions to hypothetical scenarios, but then enter the work world and act in ways that ignore ethical consequences or even violate their own values. This student might not stop to think that they have acted wrongly; or they might notice, but consider practical or professional pressures to be more important. An ethics course is only successful if it goes beyond equipping them with information and knowledge they can use but also prepares the students to scrutinize their use of that knowledge even when it is not convenient or comfortable to do so.

In order to avoid causing great harm in the world, any field

that involves practice requires not only technical proficiency of its practitioners, but also ethical proficiency, as manifested not only in a command of the relevant knowledge but also the inclination and ability to let that knowledge take precedence over laziness or self-interest. In other words, a successful professional ethics education doesn't just offer resources to indicate how problems can be identified and addressed: it also trains students to avail themselves of those resources, even when it's possible and easier not to. Teaching this to students is a complex and challenging task that cannot be successfully realized through cross-disciplinary requirements alone, but must be integrated into their CS education [55]. The number of recent professional society calls to deal with algorithmic bias and the disparate impacts of information technology systems makes clear that CS departments must engage directly with this challenging responsibility.

2.2.2 Negotiating Students' Biases

A crucial part of ethics education is helping students see beyond their own reflexive assumptions about what is true or right. Our classroom experience has shown that introducing students to three of the major schools of ethical theory — deontology, virtue ethics and utilitarianism — is an effective way to help broaden students' ability to recognize and reflect on those assumptions. While all three schools have proponents among philosophers, theologians, and other scholars who work in ethics, broader cultural discourse about ethics tends to adopt a utilitarian approach, often without any awareness that there are other ways to frame ethical inquiry. This larger cultural reliance on utilitarianism may help explain why it consistently seems, to the students, to be the most crisply-defined and "usable" of the ethical theories. But there are significant critical shortcomings to this popular version of utilitarianism. The concept of "the greatest good" is notoriously ill-defined in utilitarianism, and while trained philosophers struggle to identify or formulate a suitable definition, the gap typically goes unnoticed in less philosophical circles, enabling agents to plug in their own definition of "the good" without submitting it to scrutiny. Furthermore, it is very easy to apply the basic formula of utilitarianism, i.e., the greatest good for the greatest possible number, to a decision without a thorough consideration of all those who will be affected. This move enables agents to declare that they have pursued a morally reasoned course when, in fact, they have only calculated the benefits to themselves and those in their immediate sphere. This difficulty in attaining a sufficiently broad understanding of the effects of actions, and thus in appropriately computing the utility of those actions, can curtail the ability to have substantive ethical discussions, even insofar as everyone assents to utilitarianism.

In our experience teaching ethics courses under the auspices of computer science departments, we find that students are often drawn first to utilitarianism, perhaps because it seems more computational than the others. One of the most important aspects of the course is to broaden their experience to help them see past the non-rigorous version of utilitarianism to which they have been previously exposed. The aim is not to demonstrate the superiority of one approach over the other, but rather to help students understand the uses and limits of each approach. This limitation can be exemplified by the question of whether to replace factory workers with robots. They may focus on the happiness of the factory owners, shareholders, and those who can purchase the manufactured goods more cheaply, without considering the utility of the factory workers and those whose jobs depend on factory workers having money to spend; or even the more high level question about whether or not it is reasonable to consider human beings and machines interchangeable. Indeed, the three approaches can be complementary, or even

mutually informative; for example, recent theorists have argued that virtue ethics is best seen as part of successful deontology [38].

2.3 Why Use Fiction to Teach Ethics?

Stories — literature, plays, poetry, and other forms of narrative — have always been a way of talking about our own world, telling us what it's like and what impact our choices will have. Whether they are transmitted in print or through other media, stories play a potent role in shaping the thoughts and ideas of individuals, and the cultural norms of the societies in which they live.

Scholars of ethics have, in recent decades, embraced fiction as an ideal way to think about and teach ethics, because, as philosopher Martha Nussbaum writes, fiction “frequently places us in a position that is both like and unlike the position we occupy in life; like, in that we are emotionally involved with the characters, active with them, and aware of our incompleteness; unlike, in that we are free of the sources of distortion that frequently impede our real-life deliberations” [46]. By offering the reader both immersion and distance, an ethics course based in fiction helps students to perceive the degree to which ethical quandaries are tangled up in other aspects of life, while furnishing a context that keeps them connected to abstract principles and questions. As such, a fiction-based ethics education helps students to cultivate the capacity to recognize ethically complex situations as they arise, i.e., to extract an ethical dilemma from a larger context. This combination of qualities also helps students develop the *moral imagination* that is a key component of successful ethics education [17]. The common alternative is to provide the students with a pre-packaged case studies, e.g., the fantastic resources at <http://www.onlineethics.org/>, in which the particular ethical dilemma under study is cleanly identified for the student.

Furthermore, science fiction is particularly well-suited to teaching computer ethics. As Alec Nevala-Lee writes, “Science fiction has been closely entwined with military and technological development from the very beginning. The first true science fiction pulp, *Amazing Stories*, was founded by editor Hugo Gernsback expressly as a vehicle for educating its readers about future technology” [44]. Our project builds on this long-recognized insight: that science fiction is, in key respects, better able than “realistic” fiction to reflect the near future (or possible futures) in which computer professionals work. Science fiction, therefore, permits a curricular design that hews more closely to the concerns and quandaries of computer-related fields of study and work. A successful ethics course will reframe the task of ethical engagement so that students understand the ongoing responsibility to ask ethical questions of themselves and their work; and further, that they are equipped to perceive, describe, and understand the challenges as they arise. We claim that science fiction makes the key ethical questions of technology development and use more vivid and engaging, and the critical resources for addressing ethical questions more intelligible.

We take science fiction in its broadest sense,⁸ as the fantastical worlds or even the futuristic technology gives us a starting platform for discussion. The category of science fiction was first described by Hugo Gernsback, for whom the prestigious Hugo Prize is named, in the editorial to the first issue of *Amazing Stories* in 1926 as, “... I mean the Jules Verne, H G Wells and Edgar Allan Poe type of story — a charming romance intermingled with scientific fact and prophetic vision.” Using this broad definition, almost any fiction dealing with sufficiently advanced technology *is* science fiction. Though the majority of the literary and philosophical establishment has not, until recently, taken science fiction as a venue for

⁸Though the precise definition of Science Fiction is a matter of some debate, within the field, at the moment [58].

serious ethical thinking, this fact reflects longstanding biases in the field rather than the merits or possibilities of science fiction itself.

Fiction allows us to reframe recognizable human situations and problems in terms of unfamiliar settings and technology. Hence, any fiction, and especially science fiction in the case of technology, can be an ideal medium for raising and exploring ethical concerns. By presenting a familiar problem, e.g., conflicts between different social groups or the invasion of privacy, in unfamiliar terms and settings, a work of science fiction can curtail a reader’s tendency to defend, reflexively, their own previously-held views. As Nussbaum argues, “Since the story is not ours, we do not get caught up in the vulgar heat of our personal jealousies or angers or the sometimes blinding violence of our loves [46].” In this way, science fiction creates the opportunity for students to gain fresh insight into, and even empathy for, ethical positions and people whose real-world analogues are not embraced by their values or politics.

Hence, we advocate science fiction for several reasons:

1. The use of futuristic or alien settings allows students to detach from political preconceptions and experience the dilemmas of plot and characters as something fresh.
2. It has so far proven popular with the students.⁹ They have perceived that the course would be a chance to get credit for something they enjoy but have not found time to do while in college/graduate school: read and watch science fiction.
3. One student wrote on a Spring 2017 anonymous course evaluation, “Going into this course, there were several times that I could acknowledge an ethical situation and had my own ideas as to whether it was ‘right’ or ‘wrong’, but I couldn’t necessarily articulate why. This course gave me the tools to be able to have a meaningful discussion about these topics. It was also a productive way to get out of the coding mindset, take a step back, and consider what other results might come from the technologies that we will be making. Phenomenal course, and phenomenal instructor.”
4. By its nature, science fiction promotes discussion of possible future technology, with a focus on social implications of that technology. To quote an anonymous reviewer, “... you can discuss novel technological conundrums that may present interesting and new ethical questions.”
5. Some of the science fiction chosen also posits new science infrastructure, and allows students to think about doing research and development outside of the fairly rigid industrial and academic boxes, driven by something other than current funding paradigms. This creative thinking about practical problems, according to both philosophers [53] [41] and educators [22], is a crucial component in developing the ethical reasoning abilities of students.
5. Courses in other fields use literature (including science fiction) in non-majors courses as both a “hook” and a platform for exploring core issues [6, 21]. Scholars in other humanistic disciplines such as history and philosophy have also argued that literature is an invaluable teaching tool for ethics, among other topics (e.g., [26, 27, 49]).

⁹The initial instructor, Goldsmith, received three teaching awards the first time it was taught, and consistently received excellent teaching evaluations for it, as has Burton, who also received a teaching award based on the class.

3. THE COURSE

The aim of the course is to prepare students to recognize ethical problems in both their present and future work as technologists. To equip them for a future whose terms and conditions we in the present cannot anticipate, the course focuses on methods of applied ethical reasoning as well as on particular problems. During class discussion and in homework assignments, the students analyze both science fiction stories and brief articles, using the major ethical theories not only as evaluative tools, but as a descriptive apparatus that enable them to recognize problems and consider possible solutions from multiple perspectives. As the trajectory of many individual students' individual improvement has shown, this focus on ethical theory as a descriptive tool, combined with the use of science fiction stories as an arena for ethical description and analysis, works to sharpen students' ability to perceive and describe ethical challenges, and expands their capacity to address these challenges with both creativity and nuance. An abbreviated example syllabus is shown in Figure 1; we discuss the broad course outline in detail below.

The class opens with a crash course on ethical theories and a review of the IEEE and ACM codes of ethics. Students consider the different modes of ethical engagement invited by each code, and discuss whether, and in what ways, either code is likely to affect their decision-making. Although this discussion typically evinces varying opinions on the usefulness or relevance of either code, there is a near-universal consensus that the codes, by themselves, are not sufficient to help an IT professional address challenging problems that may arise. We the instructors stress that this is a problem common to all codes of ethics, and that the solution is not a more perfect code, but rather an IT professional better equipped to engage in ethical reasoning (and thus to make use of professional codes.)

The course then spends a few weeks on in-depth study of each of the three major ethical theories: utilitarianism, deontology, and virtue ethics. For each theory, the course spends one day on a critical reading that introduces that theory in detail, and another day analyzing and discussing a short story from within the perspective of that ethical theory. To prepare for these discussions, the students write "ethical description exercises," answering guided questions about how the story-world can be understood through that week's ethical lens. Some of these stories — particularly Elizabeth Bear's "Dolly" [7], which is used to teach deontology, and E. M. Forster's "The Machine Stops" [25], which is used to teach virtue ethics, end up becoming touchstones for the course, resurfacing in student discussions about later subjects.

After building the students' analytic competency in ethical theory, the course moves to a consideration of major ethical concerns in IT, such as surveillance, the interrelationship between news and social media, and self-driving cars. On the strength of the assigned science fiction stories, students consider both immediate practical problems and deep underlying issues that recur in IT ethics both past, present, and possibly in the future.

Each of the stories in the course touches on multiple core issues in the course, enabling the students to appreciate, and grapple with, the interconnectedness of the various challenges they will confront. Stories such as J. P. Kelly's "Itsy Bitsy Spider" [35] and Paul Shoemaker's "Today I am Paul" [52], both of which focus on carebots looking after aging parents suffering from dementia, serve as the basis for a discussion of carebots in particular, but also open out onto broader discussions of how technological interventions can change the conditions for human relationships. Paolo Bacigalupi's "The Gambler" [3] helps frame a discussion of new media and the attention economy, and highlights the particular hurdles that this new information environment creates for minority voices and posi-

tions. Ken Liu's "Here-and-Now" [36] offers a potent lens on the personal and social stakes of the post-privacy era, particularly in the context of the mostly-unregulated gig economy that depends so heavily on IT innovations. And Michael Burstein's "Teleabsence" [13] explores how technological innovations designed to address social inequality can exacerbate it, while raising probing questions about the powers and limits of how one can redefine oneself on the internet. Although the reading list has changed with each iteration, these stories and others like them have formed the backbone of each version of the course. In each case, our students have emerged from the semester's reading inspired, troubled and invigorated by the new perspectives they have gained on their future work.

The assignments in the course help develop the students' capacities for attention and critical thought in a manner intended to serve them well over the long term. By working descriptively with three different ethical theories, students develop a rich critical vocabulary for recognizing ethically fraught situations as they arise. The questions that are given to the students for a particular story are deliberately open-ended, requiring the student to identify and formulate the problem from the ground up, an approach which addresses a practical gap created when students are taught using only case studies. This open-endedness also fosters a wider range of responses than a more closely-tailored set of questions, thus creating a more varied class discussion.

Through the multiple writing assignments the students not only become aware of a range of potential ethical challenges in their work in computer science, but also become alert to the variety of ways in which these problems might initially emerge. They will be able to identify potential ethical risks in a given technology or model, or in a company's and the public's use of this technology or model. They will be prepared to articulate their arguments for why a given approach is the most (or least) ethical choice, and will also be better equipped to see past incomplete or specious defenses of potentially unethical projects.

4. EXAMPLE STORY MATERIALS

This section contains an example of the pedagogical material we have developed to help students and instructors delve into the stories they read together. The stories we have collected for this course (and, no doubt, many others we have not found) are engaging enough to spark energetic debate about ethical questions on their own. Such a story will also reward sustained scrutiny along ethical lines, but will also offer several layers of productive challenge beyond an initial encounter. Once the problems illustrated by the narrative have been described and conceptualized, the full ethical implications and challenges can be understood by "re-embedding" the problem back into its narrative contexts. The students should then consider how the world of the story created the conditions for both the external problems and internal struggles addressed by the characters.

We have developed supporting materials that are designed to capitalize on the lively accessibility of the fiction-reading experience while also helping the students come to grips with the complexity of considering a problem in the context of the wider world where it takes place. These materials include both a **story frame** to introduce the stories to students and a **pedagogy guide** to help instructors.

The story frame furnishes the students with light guidelines, preparing them to pay attention to particular issues without instructing them in how to answer, or even ask, ethical questions. The story frame thus leaves room for the student to discover the questions for themselves, and to grapple with the challenge of identifying and naming the problems at hand. This choice not only helps to pre-

	Spring 2017 Syllabus
Week 1	Using ethical language to describe / in-class discussion, read Asimov's The Dead Past [2].
Week 2	Professional ethics: Read ACM and IEEE Codes of Ethics.
Week 3	Utilitarianism: Read "Repent, Harlequin!" Said the Ticktockman [23]; Ethical description exercise #1 due in class.
Week 4	Deontology: Read Dolly [7]; Ethical description exercise #2 due in class. Virtue ethics, Read The Machine Stops [25]; Ethical description exercise #3 due in class.
	...
Week 6	Selfhood and technological mediation: Read Itsy Bitsy Spider [35], Robot carers, ethics, and older people [54]; Ethical description exercise #4 due.
	...
Week 9	Privacy: <i>Read Ken Liu's Here-and-Now</i> [36], "Privacy as Contextual Integrity" [45], M. Kakkar, "A Case against online privacy [34];" "Privacy, speech and values: what we have no business knowing" [40].
Week 14	What is Ethical Warfare? Read Codename Delphi [42], Ethical robots in warfare [1], The problem of dirty hands [24], War goes viral [12]; Ethical argument assignment #3.
Week N	Professional Ethics, the importance of integrity: Read Not Smart, Not Clever [51]; Ethical argument assignment #4. Conclusion: What can you do? What should you do? Reread The Dead Past.

Figure 1: Partial Syllabus

serve the excitement of discovery that comes with reading good fiction, but also requires the student to undertake those challenging tasks on their own. While their own initial attempts to frame, define and address the ethical problems are likely to be inadequate, their attempts to do so both individually and collectively are an essential part of the learning in an ethics course — because the real problems they encounter will also not come with a set of pre-formulated guidelines to steer them toward the best answer.

The pedagogy guide, in addition to offering generalized tips for stimulating and sustaining productive discussion about fiction and ethics, also points the instructor toward relevant themes and patterns in the text, including textual details that help indicate those patterns. *These details and patterns, by themselves, do not constitute an "answer" to any of the core ethical questions raised by the stories.* As a list of facts, they are not especially helpful for students grappling with the core ethical challenges of a given story. In the context of an ongoing discussion, the instructor can introduce these details to raise new questions, or to challenge provisional explanations, about how the world of the story works or why characters make the choices they do. In story discussions — and, indeed, in discussions of the real world — students will often begin the course by wanting to find simple, tidy answers for challenging ethical problems. To counter this impulse toward overly-neat answers, instructors will find it useful to interject into the discussion details that complicate students' tidy explanations. In this way, discussion of the story-worlds can help to train students to perceive complexity in the real world.

The story under study here is Ken Liu's "Here-and-Now," a brief story that has sparked lively and productive discussion among students in previous versions of the course. Liu, a trained computer scientist, has written several excellent stories that directly address issues in computer ethics.

4.1 Story Frame for "Here-and-Now"

This material is circulated to students along with the story text itself. "Here-and-Now" is available to read for free online at <http://www.kasmamagazine.com/here-and-now.cfm>.

"It's amazing what you can get, just by asking."
Here-and-Now

How much is information worth? That is the question that Aaron, the protagonist of Ken Liu's *Here-And-Now*, is forced to confront over the course of one complicated afternoon and evening. Aaron is one of thousands (if not millions) of people using the new app, Tilly *Here-And-Now*, which allows its users to put in anonymous requests for "information" of any and all kinds.¹⁰ This story poses deceptively simple questions about why information matters. It also points out that some kinds of information are much more meaningful or valuable to some people than to others, and asks us to consider whether that difference should matter, and how.

The world of this story is not quite the same as ours, but it is similar in many ways. It appears that Cotillion has achieved data management capabilities that have not yet been realized in our world, but the possibility is certainly on the horizon. Likewise, nothing exactly like the here-and-now app exists yet, but it is a plausible amalgam of many apps and services that do exist: TaskRabbit, Pokémon Go, and YikYak. Indeed, the app in the story was based on one described in an academic paper [57] (see also [59]). Still, we are fast approaching a world like the one in the story. It's not hard to imagine an app like this existing here, and now.

4.2 Study Questions

1. There are many essential ingredients in Tilly *Here-and-Now*'s economy: money, and information, but also interest on the part of the users: both the information-requesters and the information-gatherers. What are the sorts of interest that might lead someone to use the app, in either of those two roles? Are any of those interests in tension with one another?

2. Does it matter that Tilly's request function is anonymous? Why or why not?

3. Aaron decides, early in the story, that "Tilly *Here-and-Now* made you more aware of the world around you... more connected to your community." How do the events of the story itself confirm or challenge that conclusion? (People you can use to think about this question include Lucas, Aaron's parents, the unnamed people whose bounties are being fulfilled, the girls in the video that Lucas has purchased, and Aaron himself.)

4. The reward for fulfilling an information request is called a "bounty," instead of a "fee" or a "one-time payment" (or any other possible term — you can probably think of some others.) How does that choice of word affect the way you, the reader, think about the relationship between the information-requester and the information-gatherer? Does it affect how you think about the relationship between either of these people and the information that is gathered? Do you think the choice of the word "bounty" has an effect on the characters in the story, as well? If so, in what way?

¹⁰The story was inspired by a paper about a similar, but non-monetized, app [57]. The authors of the initial technical paper recently learned of Liu's story through a tutorial on using science fiction to teach computer ethics.

5. Who has access to the requests, and how is that access controlled?

4.3 Instructor's Guide

This material is available to instructors to help them guide in-class discussion of the story.

The Tilly Here-and-Now app exists in a world that is just different enough from ours to be provocative, but similar enough to feel intuitive. Your students may be tempted to jump straightaway to talking about the app itself, independent of the story. But this particular narrative provides an exceptionally effective window onto Liu's slightly-reimagined world, and your discussion will likely be more focused and productive if you dedicate at least 20–30 minutes to talking about Aaron's experiences and reflections before moving onto the more general implications of the Tilly app.

As always, the best approach will be a Socratic one, in which you guide the students toward discovering things for themselves. Below are some observations and details about the story. You can use these to ask “fishing” questions if you think the students are missing important details, or to prompt them to reassess their view of the story if they have settled on a version that ignores such details.

Aaron. Aaron is really interested in information about others. He likes claiming bounties and furnishing others with the information they want, but he also likes trying to figure out why it is that people want them. When Lucas baits him by saying “I got something cool,” Aaron can’t help asking about it: he wants to know the answer. On the other hand, Aaron really hates giving up information about himself to the people he knows. He doesn’t want his mother to know about his part in the play, and he won’t tell Lucas how much money he has earned. In the entire story, we only learn of one instance in which Aaron has willingly imparted information to someone else: when he teaches his mother (before the beginning of the story) about Tilly Here-and-Now. And by the end, Aaron regrets having shared that information, since his mother is now using the app “against” him, to learn things about him and about his father.

The Individuality of Knowledge. At several points, both major and minor, the story directs our attention to the ways in which information matters more to the people whom it touches directly. In this way, the story adds a new layer to frequently-expressed concerns about privacy, which focus on the damage that is done to the person(s) whose information is known or made available. According to Liu, the person who knows can be just as affected or damaged by the knowledge as the subjects about whom it is known.

Lucas is happy with his video of two girls kissing (which strikes Aaron as invasive of the girls’ privacy), but “Would have been even better if they’re people I know,” Lucas said. “Next time I’m going to raise the bounty and limit the range more. It’s amazing what you can get just by asking. [36]” Lucas anticipates that knowing the girls involved would make the video more satisfying. The invasion of private space is part of the pleasure.

Aaron and Lucas respond very differently to the license plate request: not only because Aaron recognizes the plate number, but because he has something personal at stake in the fulfillment of the request — and in its asking. Before this, we have seen Aaron wonder why he is being asked to fulfill this or that request, but never wondering whether he should. Only when the request touches him personally does he realize the damage that might be done if it is fulfilled.

Aaron himself is later undermined (in a small way) by another Tilly Here-and-Now user, fulfilling another of his mother’s requests, when she discovers that he has been cast in the play. On the surface, this plot point lines up primarily with more typical concerns about privacy. Aaron, who had hoped to conceal the information, is the one who has been injured, but insofar as Aaron trusts his mother less, she is also damaged.

Information Control and Performance. The story repeatedly touches on the theme of people pretending to be who they are not. This theme is signaled at the opening of the story, when we learn that Aaron has been cast in a play. A play is a performance, but the “deception” is a matter of mutual consent: the audience knows it is watching actors, and in this sense a play does not represent a miscarriage of knowledge.

This non-deceptive deception differs from the way that Aaron’s parents treat each other over dinner toward the end of the story. Aaron knows by then that his mother suspects his father of cheating, and he halfway suspects his father, as well, but they treat each other normally, as if nothing were wrong. “He couldn’t hear anything different in their tones. His mother acted like she had never asked the question. His father acted like he had nothing to hide [36].” Aaron’s mother, and possibly his father as well, are performing with an intent to mislead. But whom are they misleading: Aaron, or each other, or both? And when did the deception begin?

It is also worth raising for discussion the question of whether, and how, Aaron’s own actions qualify as misrepresentations: his desire to keep the play role a secret, and his own Tilly request, which is designed to distract Lucas from fulfilling his mother’s request.

Additional Topics. This story also raises issues of access control, and of information integration — that is, combining different possibly-innocuous sources to complete more complex, thorough, and possibly invasive records. One of the moments in which a reader’s perception shifts is that moment when Aaron recognizes his father’s license plate. How do the different ethical theories frame the possibility of deanonymization, either deliberate or accidental? This can lead to discussion of hacking/Wikileaks, or to trust in data scrubbing, among other directions.

4.4 Ethical Description Writing Assignment

The purpose of this assignment is description. Addressing the points listed below, your task is to describe Liu’s story in terms of one of the three major theories of ethics. (You will receive separate instructions telling you which theory to use.) Be sure to title your assignment “Here-and-Now: [name of ethical theory]”

- *Using the concepts and worldview of your assigned theory, give a 2–4 sentence summary of the central ethical problem(s) in the story.*
- *What is at stake in the ethical problem(s), so described? (That is: what possible goods could be gained/lost, or what kinds of harm could occur/be prevented?) Using the language of your theory, explain why these costs or benefits are significant.*
- *Which character(s) is/are in a position to take meaningful action, with respect to that problem? What about their character or circumstances positions them to take meaningful action?*
- *Choose one character from your answer above. Using the language and concepts of your assigned theory, describe the course of action that this character takes in the story.*
 - *Are there other possible courses of action that the story suggests the character might have taken? (Answer: probably.) Describe those alternate courses of action, again using the language of your assigned theory.*
 - *According to that theory, which is a better course of action, and why?*

- *What argument do you think the ending of the story is making? (You are still describing here, rather than arguing; your job is to use the language of your assigned theory to describe Liu's argument.)*

You will bring these assignments with you to class on the day they are due. You are welcome to make notes on them, over the course of discussion, for your own edification, since ultimately you will get to keep them. You will turn them into me at the end of class. They will be returned with a grade and comments by the next class.

5. CONCLUSION

Teaching ethics to computer science students is a pressing responsibility for computer science faculty, but also a challenging one. This article argues that using fiction as the basis for an ethics course offers several advantages beyond its immediate appeal to many students and some faculty. First, fiction offers the students a way to engage with ethical questions that helps them cultivate their capacity for moral imagination; science fiction in particular can make the ethical stakes of blue-sky projects feel pressing and immediate. Second, stories offer the students the chance to develop their skills in ethical description. Finally, discussing ethics in the context of fiction can make it easier for instructors to adopt an open-ended approach that is required for a good ethics course.

A course built around fiction enables instructors to incorporate the best and most useful aspects of a humanistic approach to ethics education, while remaining closely geared to the concerns of computer science. To help clarify our approach, and to demonstrate how it can be used by instructors with no prior background in ethics, we have provided teaching materials for one publicly available story.

Acknowledgments

We wish to thank the anonymous reviewers for very helpful feedback. We would also like to thank Cory Siler and John Fike for proofreading, and for acting as sounding boards. This material is based upon work supported by the National Science Foundation under Grant No. 1646887. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

6. REFERENCES

- [1] ARKIN, R. C. Ethical robots in warfare. *IEEE Technology and Society Magazine* 28, 1 (2009), 30–33.
- [2] ASIMOV, I. The dead past. *Astounding Science Fiction* (1956). Available in more recent collections.
- [3] BACIGALUPI, P. The Gambler. In *Fast Forward 2*, L. Anders, Ed. Pyr, 2008.
- [4] BATES, R., GOLDSMITH, J., BERNE, R., SUMMET, V., AND VEILLEUX, N. Science fiction in computer science education. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (2012), ACM, pp. 161–162.
- [5] BATES, R., GOLDSMITH, J., SUMMET, V., AND VEILLEUX, N. Using science fiction in CS courses. In *Proceedings of the 45th ACM technical symposium on Computer Science Education* (2014), ACM, pp. 736–737.
- [6] BATES, R. A. AC 2011-1669: AI & SciFi: Teaching writing, history, technology, literature and ethics. In *ASEE Annual Conference & Exposition* (2011).
- [7] BEAR, E. Dolly. *Apex Magazine* (2016).
- [8] BEARD, M. Ethics, morality, law — what's the difference? <http://www.ethics.org.au/on-ethics/blog/september-2016/ethics-morality-law-whats-the-difference>, September 2016. Online; posted 27-August-2012.
- [9] BONNEFON, J.-F., SHARIFF, A., AND RAHWAN, I. The social dilemma of autonomous vehicles. *Science* 352, 6293 (2016), 1573–1576.
- [10] BOYER, E. L. The scholarship of engagement. *Journal of Public Outreach* 1, 1 (1996), 11–20.
- [11] BOYER, E. L. *Scholarship Reconsidered: Priorities of the Professoriate*. Jossey-Bass, 1997.
- [12] BROOKING, E. T., AND SINGER, P. War goes viral: How social media is being weaponized across the world. *The Atlantic* (2016).
- [13] BURSTEIN, M. A. TeleAbsence. In *I Remember the Future*. Apex Publications, 2008.
- [14] BURTON, E., GOLDSMITH, J., KOENIG, S., KUIPERS, B., MATTEI, N., AND WALSH, T. Ethical considerations in artificial intelligence courses. *arXiv preprint arXiv:1701.07769* (2017). Extended version of *AI Magazine* article of the same name, Summer 2017.
- [15] BURTON, E., GOLDSMITH, J., AND MATTEI, N. Teaching AI ethics using science fiction. In *1st International Workshop on AI, Ethics and Society, Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence* (2015).
- [16] BURTON, E., GOLDSMITH, J., AND MATTEI, N. Using “The Machine Stops” for teaching ethics in artificial intelligence and computer science. In *1st International Workshop on AI, Ethics and Society, Workshops at the Thirtieth AAAI Conference on Artificial Intelligence* (2016).
- [17] CALLAHAN, D. Goals in the teaching of ethics. In *Ethics Teaching in Higher Education*. Plenum Press, New York, 1980, ch. 2, pp. 61–80.
- [18] COLBECK, C. L. Merging in a seamless blend: How faculty integrate teaching and research. *The Journal of Higher Education* 69, 6 (1998), 647–671.
- [19] COLBY, A., AND SULLIVAN, W. M. Ethics teaching in undergraduate engineering education. *Journal of Engineering Education* 97, 3 (July 2008), 327–338.
- [20] DAVIS, B. G. *Tools for Teaching*, second ed. Jossey-Bass, 2009.
- [21] DILS, L. S. Science fiction and the future, Yale-New Haven Teachers Institute, 1987. <http://www.yale.edu/ynhti/curriculum/units/1987/2/87.02.04.x.html>.
- [22] EDMISTON, B. Ethical imagination: Choosing an ethical self in drama. In *Imagining to Learn*. Heinemann, 1998.
- [23] ELLISON, H. “Repent, Harlequin!” Said the Ticktockman. Underwood Books, December 1965.
- [24] ELSHTAIN, J. B. Reflection on the problem of “dirty hands”. *Torture: A collection* 83 (2004).
- [25] FORSTER, E. The machine stops. *The Oxford and Cambridge Review* (November 1909). Multiply republished.
- [26] GARCIA IOMMI, L. Let's Watch a Movie!: Using film and film theory to teach theories of international politics from a critical perspective. *American Political Science Association 2011 Annual Meeting* (2011).
- [27] GOERING, S. Using children's literature as a spark for philosophical discussion: Stories that deal with death. In

Ethics and Children's Literature, C. Mills, Ed. Ashgate, 2014.

[28] GOLDSMITH, J., AND BURTON, E. Why teaching ethics to AI practitioners is important. In *Proceedings of the Thirtyfirst AAAI Conference on Artificial Intelligence* (2017).

[29] GOLDSMITH, J., AND MATTEI, N. Science fiction as an introduction to AI research. In *Proc. 2nd AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI)* (2011).

[30] GOLDSMITH, J., AND MATTEI, N. Fiction as an introduction to computer science research. *ACM Transactions on Computer Science Education* 14, 1 (Mar. 2014), 4:1–4:14.

[31] GREENFIELD, A. *Radical Technologies: The Design of Everyday Life*. Verso, 2017.

[32] HAWORTH, J. G., AND CONRAD, C. F. Curricular transformations: Traditional and emerging voices in the academy. In *Revisioning Curriculum in Higher Education*, J. G. Haworth and C. F. Conrad, Eds. Simon & Schuster Custom Publishing, 1995, pp. 191 – 202.

[33] HUFF, C., AND FURCHERT, A. Toward a pedagogy of ethical practice. *Commun. ACM* 57, 7 (July 2014), 25–27.

[34] KAKKAR, M. A case against online privacy, 2011. *ZDNet Online*, [Online; accessed 4-October-2016]. <http://www.zdnet.com/article/a-case-against-online-privacy/>.

[35] KELLY, J. P. Itsy bitsy spider. *Asimov's* (June 1997).

[36] LIU, K. Here-and-now. *Kasma* (2013). 1 Nov.

[37] MARTIN, C. D., AND WELTZ, E. Y. From awareness to action: integrating ethics and social responsibility into the computer science curriculum. *ACM SIGCAS Computers and Society* 29, 2 (June 1999), 6–14.

[38] MCNAUGHTON, D., AND RAWLING, P. Deontology. In *The Oxford Handbook of Ethical Theory*, D. Copp, Ed. Oxford University Press, 2006.

[39] MIHAIL, R., RUBIN, B., AND GOLDSMITH, J. Online discussions: Improving education in CS? In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education (SIGCSE 2014)* (2014).

[40] MOORE, A. D. Privacy, speech, and values: what we have no business knowing. *Ethics and Information Technology* 18, 1 (2016), 41–49.

[41] MURDOCH, I. The idea of perfection. In *The Sovereignty of Good*. Routledge & Kegan Paul, 1970.

[42] NAGATA, L. Codename: Delphi. *Lightspeed* 47 (April 2014).

[43] NARAYANAN, A., AND VALLOR, S. Why software engineering courses should include ethics coverage. *Communications of the ACM* 57, 3 (Mar. 2014), 23–25.

[44] NEVALA-LEE, A., 2016. WorldCon 2017 presentation; personal communication 1/15/2017.

[45] NISSENBAUM, H. Privacy as contextual integrity. *Washington Law Review* 79 (2004), 119.

[46] NUSSBAUM, M. *Love's Knowledge: Essays on Philosophy and Literature*. Oxford University Press, 1990.

[47] PEASE, A. Teaching ethics with science fiction: A case study syllabus. *Teaching Ethics: The Journal of the Society for Ethics Across the Curriculum* 9 (2009), 75–82.

[48] PERRY, W. G. Cognitive and ethical growth: The making of meaning. In *The Modern American College*, A. W. Chickering and Associates, Eds. Josey Bass, 1980, pp. 76–109.

[49] RODWELL, G. *Whose History?: Engaging History Students Through Historical Fiction*. University of Adelaide Press, 2013.

[50] ROGAWAY, P. The moral character of cryptographic work. *Cryptology ePrint Archive*, Report 2015/1162, 2015. <http://eprint.iacr.org/>.

[51] SAXEY, E. Not smart, not clever. *Apex Magazine* (2014).

[52] SHOEMAKER, M. L. Today I am Paul. *Clarkesworld* (August 2015).

[53] SMITH, B. Analogy in moral deliberation: the role of imagination and theory in ethics. *Journal of Medical Ethics* 28, 3 (Aug. 2002), 244–248.

[54] SORELL, T., AND DRAPER, H. Robot carers, ethics, and older people. *Ethics and Information Technology* 16, 3 (2014), 183–195.

[55] SPRADLING, C. L. *A study of social and professional ethics in undergraduate computer science programs: Faculty perspectives*. PhD thesis, The University of Nebraska-Lincoln, 2007.

[56] THE EXECUTIVE BOARD OF THE AMERICAN ANTHROPOLOGICAL ASSOCIATION. Statement on human rights. *American Anthropologist* 49, 4 (1947), 539–543.

[57] VON DER WETH, C., AND HAUSWIRTH, M. Finding information through integrated ad-hoc socializing in the virtual and physical world. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (2013), vol. 1, IEEE, pp. 37–44.

[58] WALLACE, A. Who won science fiction's Hugo awards, and why it matters, 2015. *Wired Magazine*, [Online; accessed 22-October-2015]. <http://www.wired.com/2015/08/won-science-fictions-hugo-awards-matters/>.

[59] WETH, C. V. D., ABDUL, A. M., AND KANKANHALLI, M. Cyber-physical social networks. *ACM Transactions on Internet Technology (TOIT)* 17, 2 (2017), 17.