Designing Ethically-Integrated Assignments: It's Harder Than it Looks

Noelle Brown[*]
University of Utah
Salt Lake City, UT, USA
noelle.brown@utah.edu

Koriann South[*]
University of Utah
Salt Lake City, UT, USA
korianns@ymail.com

Suresh Venkatasubramanian
Brown University
Providence, RI, USA
suresh_venkatasubramanian@brown.edu

Eliane S. Wiese
University of Utah
Salt Lake City, UT, USA
eliane.wiese@utah.edu

ABSTRACT

While the CS education community has successfully incorporated tech-ethics assignments and modules into computing courses, we lack a defined process for instructional design to create these materials from scratch across the curriculum. To enable the development of such a process, we explore two research questions: (1) What specific instructional design challenges emerge when creating ethically-integrated assignments for CS courses? And (2) what strategies might overcome them? We address these questions using Research through Design, a method for critically examining design processes. Applying this method to our own process of creating ethics-integrated CS assignments yielded four key challenges: identifying an ethical context, maintaining a technical focus, eliciting both ethical and technical thinking from students, and making the assignment practical for the classroom. Further, the Research through Design approach revealed process-level insights for addressing these challenges, which can apply across the computing curriculum. This paper also serves as a case study of Research through Design for CS education, highlighting the importance of the instructional design process and the behind-the-scenes challenges and design decisions that go into tech-ethics materials.

CCS CONCEPTS

• **Social and professional topics → Computing education**.

KEYWORDS

ethics, responsible computing, instructional design, research through design

ACM Reference Format:

Noelle Brown, Koriann South, Suresh Venkatasubramanian, and Eliane S. Wiese. 2023. Designing Ethically-Integrated Assignments: It's Harder Than it Looks. In *Proceedings of the 2023 ACM Conference on International Computing Education Research V.1 (ICER '23 V1), August 07–11, 2023, Chicago, IL, USA*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3568813.3600126

[*]Both authors contributed equally to this research.

ICER '23 V1, August 07–11, 2023, Chicago, IL, USA

ACM ISBN 978-1-4503-9976-0/23/08.
https://doi.org/10.1145/3568813.3600126

1 INTRODUCTION

In response to many high-profile cases of technology posing clear ethical concerns (e.g., autonomous cars [7], facial recognition [9], criminal justice [44]), researchers and educators have urged computing programs to integrate ethics within Computer Science (CS)[1] courses [21, 28]. However, from an instructional design perspective, it's not clear how to incorporate ethics across the CS curriculum. Indeed, a recent survey of CS instructors by Smith et al. [64] found that, of 138 respondents, 51% thought that some topics in computing did not involve any ethical considerations - this was the reason that many instructors gave for not including ethics in courses that focused on abstract, mathematical topics [64]. Yet, instructors overall supported the general idea of integrating ethics into CS courses [64]. Together, these results suggest an instructional design gap: instructors want to incorporate ethics in their courses, but often aren't sure how, or if it's even possible. Sharing instructional resources is a useful way to support other instructors, who can reuse the materials directly, modify them, or build off them as inspiration. Instructors cited these resources as a helpful asset for integrating ethics within introductory courses [64]. However, exemplar materials were less useful for instructors of advanced or niche courses, who did not find their course content reflected in the material, and could not easily generalize from the examples to create instruction suited to their topics [64].

Our work responds to the call from Smith et al. [64] to generate ethics assignments for a variety of computing courses. However, rather than creating sample assignments for particular courses, we aim to support instructors in creating their own materials, for any course. To that end, we argue that our community should share not only the finalized tech-ethics lessons and assignments, but also the thought processes, decisions, and challenges that went into creating those materials. That is, we need to share the process, not just the outcome. By examining our instructional design processes, we can identify useful steps that others can follow, regardless of topic. To be clear, we did not set out to explore an instructional design process for incorporating ethics into CS. At first, we simply wanted to create ethically-integrated assignments for an Artificial Intelligence (AI)

[1]In this paper, we use the terms 'computing' and 'CS' interchangeably to refer to the field of computing.

course. We did not anticipate significant challenges in the design process: the instructor was supportive, and AI was a natural fit for ethics content. To our surprise, creating even one assignment was very challenging, and, to our knowledge, the nuances of the obstacles we encountered had not been documented in prior work. This motivated our research questions:

RQ1 What are the instructional design challenges in creating ethically-integrated assignments for CS courses?
RQ2 What strategies can help overcome these challenges?

Since these questions pertain to the instructional design process itself, we answer them by following a Research through Design (RtD) methodology [75]. RtD, a design research method from the field of Human-Computer Interaction (HCI), enabled us to identify design challenges and suggest methods to overcome them. The challenges span the entire instructional design process, from brainstorming how ethics may relate to a specific technical topic, to implementing these ideas such that they are realistic for a classroom assignment, to ensuring the assignment is assessable and feasible for the course staff to grade.

We present a design contribution: specific design challenges and lessons learned for instructors to consider when designing assignments that integrate ethics yet remain congruent with the technical learning outcomes of their course. Specifically, the contributions of this paper are twofold: (1) identified challenges in designing ethically-integrated assignments for a technical course and (2) a presentation of suggestions to overcome these challenges that can be broadly applied to instructional design for ethically-integrated assignments in CS courses.

In the sections that follow, we orient our work with a description of our approach to ethics integration in CS courses followed by a chronological description of the instructional design process that we followed to integrate ethics within a technical assignment. We frame our design process with the overarching challenge related to each phase along with design considerations that instructors can use to overcome these challenges when creating their own ethically-integrated technical assignments. Our design case explains the decisions we made to integrate ethics into a technical assignment on graph search algorithms and our evaluation of the assignment with three think-alouds. We also describe the design considerations an instructor used to modify this assignment to be feasible for 120 students in an undergraduate course.

2 BACKGROUND

Many researchers and educators acknowledge the importance of preparing students to recognize the ethical implications of their technological decisions [22, 25]. However, designing ethical instruction for technical courses can be difficult, particularly when there are no apparent connections between the course content and ethics [64]. Additionally, there are numerous ways to teach tech-ethics (e.g., discussing the creation or use of technology [40], evaluating technology through ethical frameworks [10], or analyzing social impacts [11]), making it challenging for instructors to know how they should do this in their own courses. To support instructors for all courses in this effort, we focus on instructional design methods that can be broadly applied across the curriculum. In the remainder of this section, we describe our approach to teaching ethics in CS (Sections 2.1 and 2.2), followed by background on the design research methodology we followed to answer our research questions (Section 2.3).

2.1 What do we Mean by "Ethics"?

In CS education, teaching "ethics" can encompass various aspects, ranging from formal ethical frameworks [10, 59] to decision making [2] to social justice [45]. In our work, we draw on three types of ethics: personal, philosophical, and professional ethics, as described by Davis [17]:

- **Personal ethics**: general standards of conduct, like alerting a stranger that they dropped their wallet.
- **Philosophical ethics**: determining duty, responsibility, and rights, and the reasoning that undergirds those claims.
- **Professional ethics**: standards of conduct for domain experts within their fields (e.g., research ethics, engineering ethics, medical ethics, etc.).

While computing domain experts may be obligated to follow ethical behavior in daily life (which falls under the category of personal ethics), they also have additional ethical standards based on the duties of their profession [3]. These special obligations come from the nature of contexts that require both technical and ethical expertise. Our focus is on professional ethics, where CS expertise is required to make a sound ethical decision in such contexts. For example:

- A philosophy expert can explain the reasoning for environmentally sensitive computing, but only a CS expert can analyze algorithms to determine which are less energy-intensive.
- A layperson can advocate for algorithmic transparency, but only a CS expert can determine if the human-readable representation is accurate.

We are concerned with supporting the teaching of professional ethics, contexts where CS expertise is required to make a decision with potential helpful or harmful impacts.

2.2 Designing Assignments to Teach Professional Ethics in CS

One approach to teaching professional ethics in computing courses is to contextualize technical assignments using ethical narratives (e.g., [18, 21, 23, 34, 48]). For example, a coding assignment on loops and lists can be contextualized to require students to use these technical skills to design and code a hiring algorithm [18]. The contextualized assignment presents ethical questions including algorithmic fairness, while maintaining the goal of having students practice loops and lists [18]. This method aligns ethical components with technical learning goals [72], reveals human judgments in technical contexts, and highlights the social consequences of technical decisions. Such assignments help instructors to incorporate ethics while assessing technical material.

While we do not claim that this approach teaches ethics in its entirety [45, 63] or guarantees ethical behavior from students [30], it can help students recognize that they are making human judgments in their work. We argue that the first step for students to develop their ethical reasoning is to recognize that an ethical issue exists [19]. Once students realize that there exist human judgments that they are making in their work, consciously or unconsciously, they can begin to ethically reason through their choices, relying on the professional opinions of experts in ethics where needed. Our goal is to design assignments that help students with this first step. We want students to recognize that they need to decide if a technical decision is appropriate and when a technique needs to be modified.

We have a growing body of example assignments that give students practice applying ethical reasoning to their technical work (e.g., [18, 21, 23, 48, 55, 61]), but it can be challenging to design these assignments for our own technical learning goals. A formal discussion of design guidelines for such assignments would be a beneficial support, particularly for advanced or niche computing courses without existing examples or online resources [64].

2.3 Research through Design

To discuss design guidelines, we need instructional design research. Research through Design (RtD) is a methodology that bridges the gap between research and design by maintaining that the design process itself (i.e., making, discussing, reflecting, and iterating) can generate new knowledge [75]. What characterizes this approach as a research process rather than a design practice is that it is more explicitly reflective, placing an emphasis on the process of interpreting and reinterpreting design choices and rationale [75]. Zimmerman and Forlizzi [75] suggest five steps for conducting an RtD research project:

(1) **Select**: Choose a research problem.
(2) **Design**: Iteratively create a new artifact while documenting design decisions and rationale.
(3) **Evaluate**: Evaluate the artifact.
(4) **Reflect and Disseminate**: Reflect on knowledge gained throughout the design process and disseminate the research through publication.
(5) **Repeat**: Repeat this process.

RtD provides a solution to a challenge in the field of HCI that design products must come before theory (e.g., researchers could only evaluate the usefulness of the computer mouse after it was designed) [12]. Applied to educational research, we can only prove the usefulness of instructional design after its creation [57, 60]. Using an RtD approach, we contribute design research that can later be empirically evaluated. This reflective process allows us to formally identify the challenges and suggestions in designing ethically-integrated technical assignments. We expand on how we applied RtD in Section 3, responding to the call to action from Zegura et al. [74] for HCI scholars to contribute to the study and design of computing ethics education. To make this work accessible to both researchers and instructors, we present our results in the style of translational research (Section 4) [49], in which we translate our research concepts to classroom practice (in the form of actionable design suggestions) with the hope that instructors can immediately apply our design insights in their classrooms [39].

3 METHODS

At a high level, we aimed to identify design considerations for ethically-integrated assignments by cataloging issues encountered during the iterative revision process. Our approach followed the steps outlined by Zimmerman and Forlizzi [75] for conducting RtD research projects (discussed in Section 2.3).

3.1 Select: Defining Research Questions

Our research aimed to examine the difficulties of developing ethics-integrated assignments for CS courses and translate what we learned into design recommendations for instructors and instructional designers. To accomplish this, we formulated the following research questions:

RQ1 What are the instructional design challenges in creating ethically-integrated assignments for CS courses?
RQ2 What strategies can help overcome these challenges?

3.2 Design: Understanding the Instructor's Needs and Iterative Design

To ensure the practicality of our work, we consulted the AI instructor at our university to identify opportunities for integrating ethics. The assignments for his course were based on material from the Introduction to AI course (CS 188) at the University of California, Berkeley [66]. Materials for CS 188 (including lectures, assignments, schedule, and instructor guide) are freely available, allowing this course to be a model for many institutions, including our university. Wiese (an expert on CS education) and South selected two assignments on graph search algorithms (Appendix A) and Bayesian probability, which lacked real-world or ethical contexts, and worked together for 12 weeks to integrate ethical issues into these assignments (Appendix B and D). Wiese and South recorded notes of their discussions and reflected on the design process during each weekly meeting. Venkatasubramanian (an expert in AI ethics), an additional expert in security, and the AI instructor reviewed the assignments during week four and after week 12, and Wiese and South incorporated their feedback into the revision process. This collaboration ensured that the tasks aligned with both AI and ethics learning goals and were appropriate for the class.

3.3 Evaluate: Think-Aloud Protocol

To evaluate our assignments, we conducted think-alouds with three participants recruited from the previous semester's AI class at our university. Participants were compensated $15 for a one-hour audio-recorded session, approved by our IRB (protocol #00136739). South shared the online assignments with participants and recorded their screens as they worked through the tasks. We then transcribed the recordings, and Brown applied an inductive coding approach, focusing on capturing students' utilization of AI content knowledge and ethical considerations while also noting moments that were deemed confusing, emotional, or exhibited particularly positive or negative responses. The codes were further refined and consolidated in collaboration with South [47]. To ensure a comprehensive analysis, both coders independently re-coded the transcripts and collaboratively organized the codes into overarching themes. This process provided us with valuable insights into students' problem-solving processes, areas of confusion, and ethical considerations, which enabled us to explore commonalities and variations in student responses to assess the effectiveness of our assignments in eliciting simultaneous technical and ethical thinking. Results from the think-alouds are presented alongside identified challenges to provide concrete data. Participant responses are presented verbatim, but filler words (e.g., "like," "um") are removed for clarity. Wherever participant responses are included, it is important to note that we are discussing how to evaluate the assignment, not how to assess the students, and it was made clear to participants that we were testing our assignment, not evaluating their knowledge.

3.4 Reflect and Disseminate: Explicit Reflection

By thoroughly reviewing our meeting notes and engaging in explicit reflection on our design process and the obstacles we faced, we distilled our insights into four primary design challenges. Initially, we compiled a comprehensive list of encountered roadblocks, and through careful consideration of the underlying themes, we grouped them into broader challenges. Through iterative refinement, we consolidated similar challenges until we had a comprehensive representation of the significant pain points, culminating in four overarching design challenges. We proceeded to align the themes identified in the think-aloud evaluation (Step 3, Section 3.3) with our four challenges, undertaking a thoughtful examination of what we could have approached differently in each challenge, drawing from what we learned throughout our design process and the participants' interpretations of our materials or understanding of the content. By synthesizing these lessons, we derived design suggestions tailored to creating ethics-integrated assignments, detailed in Section 4. Although we created two such assignments (graph search algorithms and Bayesian probability), this work primarily focuses on the former to exemplify our considerations, with some support from the latter. Nonetheless, the design suggestions were developed by reflecting on the design process for both assignments, and we believe our suggestions are transferable to designing ethically-integrated assignments for various courses.

3.5 Repeat: Qualitative Analysis and Additional Reflection

The AI instructor at our university adapted the graph search assignment from our RtD project to fit his classroom and learning goals (repeating Step 2, Section 3.2). The assignment (Appendix C) was given to 120 undergraduate AI students and was qualitatively analyzed for validation (repeating Step 3, Section 3.3). Our prior work details this analysis [8]. To enhance our design research, we describe the instructor's experience in modifying the assignment for his classroom (repeating Step 4, Section 3.4), offering valuable lessons and insights for other instructors in designing their own assignments.

3.6 Positionality

The RtD process requires us to examine our own design process; therefore, our positionality inherently shapes our perspective and methodology. We approach this research with a commitment to reflexivity, acknowledging that our backgrounds, experiences, and identities impact our work. In particular, conceptions of ethics vary across cultures – our view of ethics broadly, and the type of ethics we explore in this specific assignment, reflect our context as academics in a WEIRD society (Western, educated, industrialized, rich, and democratic) [33]. Since each researcher's background individually shapes their contribution to the design process, we present personal positionality statements written by each author.

Brown: During my master's degree in data science, I was lucky to learn from particular professors who emphasized the importance of ethics in the field, and I tried to do the same when I started teaching data science myself. This was tougher than I thought, and those challenges motivate my research in ethics education. However, my thinking about ethics and its integration is inherently limited and shaped by my background and identity as a white woman raised in the U.S. I joined this project after the assignment design and think-alouds were completed, allowing me to analyze my co-authors' meeting notes and design challenges from an outside perspective.

South: I'm a person of mixed race who identifies predominantly as Black. My well-being and that of my BIPOC family and friends are obviously important to me, and I hold a healthy, historically grounded, and I hope understandable apprehension toward new technologies (e.g., COMPAS's recidivism prediction [44], recognition software unable to recognize darker skin tones [51], ChatGPT snafus [4], etc.). My objective in this research is to assist instructors in guiding future technologists to create more equitable technologies. As the facilitator of the think-aloud sessions, an undergraduate student, and a Black person who has participated in many awkward race-related conversations, I am aware of the challenges involved in participating, even as an observer, in discussions related to inequality. As a developer of educational tech-ethics materials, I also understand the difficulty in anticipating how one's words, assignments, and other actions may be interpreted. In Section 4.4.2, we discuss how I did not anticipate the interviewee making a mathematical error that resulted in a statement that could be seen as discriminatory. It's not easy to facilitate or participate in discussions about tech-ethics issues. But as Computer Scientists (instructors, students, or otherwise), I think we have a responsibility to try.

Venkatasubramanian: My background as an Indian male and an immigrant influence my thinking and my research. As a brown-skinned minority in the US (who is often identified as the "model" minority [73]), and a majority representative within computer science, I am simultaneously aware of my privilege within my area of academic expertise, the privilege accorded to people with my background, but also how we are set against other minorities as a way to preserve structures of bias and discrimination. This fuels my work in algorithmic bias and socially responsible technology: I feel the urgency of using the platform that I am afforded, and the rooms I am allowed into, to speak for those who aren't given the same opportunities.

Wiese: I am a Jewish, white-passing cis-gendered woman. My approach to education research is grounded in cognitive aspects of learning [43], instructional design processes that emphasize the alignment of goals, instruction, and assessment [13], and knowledge integration [46]. My teachers included the creators of Research through Design. I explore how to teach ethics in computing because of Ko's inspiring presentation of it as a grand challenge [41]. As a Jew, I am sensitive to the social construction of race and ethnicity, and the importance of historical context in examining socio-technical issues. Broadly, my research aims to support students in making connections between computing, ethics, and inclusivity within their social contexts.

4 DESIGN CHALLENGES

In this section, we present our results from the RtD process in the form of four main design challenges and recommendations to overcome them:

- **Challenge 1**: Identifying an ethical context. **Suggestion**: Search for nuances and student misconceptions related to the technical learning goals, then choose an ethical context that highlights human judgements involved in the technical implementation.
- **Challenge 2**: Integrating ethical components while maintaining a technical focus. **Suggestion**: Ensure that questions require both technical and ethical reasoning by examining the knowledge required for each task, and striving for increasing technical depth where students must change their technical answers based on their ethical decision making.
- **Challenge 3**: Helping students integrate technical and ethical thinking. **Suggestion**: Anticipate students' tendency to answer questions from either a purely technical or purely philosophical perspective, and write clear instructions that explicitly encourage them to use both technical and ethical modes of reasoning to answer each question.
- **Challenge 4**: Making the assignment practical for the classroom. **Suggestions**: (1) Scale the assignment using autogradable questions, and (2) prepare for errors in students' technical execution that could result in incorrect or harmful reasoning about the broader ethical issues.

We describe our design steps and our thinking at the time, interpreted through the lens of our later reflections. Where appropriate, we provide quotations from the think-alouds to illustrate these challenges. The challenges summarize the main difficulties we faced at each stage of the design process and are presented chronologically in the order of the design process that we followed.

4.1 Challenge 1: Identifying an Ethical Context

Suggestion: Search for nuances and student misconceptions related to the technical learning goals, then choose an ethical context that highlights human judgements involved in the technical implementation.

To find assignments where we might incorporate ethics, we reviewed publicly-available problem sets from an AI course [66] and tried to match the technical content with real-world incidents. Given the volume of news stories related to AI systems with ethical consequences (e.g., [44, 68, 71]), we thought this process would be simple. It wasn't. We could not find ethical contexts for abstract topics, including Minimax Theory, Markov Decision Processes, Reinforcement Learning, and Hidden Markov Models. We realized that attention-grabbing ethical controversies – e.g., predictive policing, facial recognition, and rankings for search results [52] – were not applicable to the algorithms in this class. Further, many of the course algorithms did not involve training data, making entire categories of ethical data issues irrelevant (e.g., permissions, bias, and feedback loops). Although we consulted a list of real-world contexts [52] to match with assignments, we could not use any of those examples directly because the algorithmic issues were not aligned to our course.

We found an ethical context for graph search algorithms (e.g., Breadth-First Search, Depth-First Search, A^* Search), because they reminded Venkatasubramanian of a real-life controversy in Boston [6]. School administrators tried to save money by algorithmically optimizing the school bus routes, but the algorithm ended up amplifying systemic bias, reserving the most favorable bus routes for affluent neighborhoods [29]. While we did not use this busing context, the example highlighted how the technical decisions developers may take for granted, such as the choice of optimization function, can have ethical consequences.

We started by looking for real-world examples of ethical issues related to the target algorithms for two main reasons: (1) we wanted authentic contexts, and (2) we couldn't easily imagine human contexts for these algorithms that could involve ethical problems. Looking back, our approach of trying to start from a human context was narrow minded. We made more progress by starting from the technical learning goals of the assignment, and considering what problems might result if the student applied their knowledge by rote (that is, when implementation choices are made implicitly, following defaults, rather than explicitly, in response to context).

This approach ensures that the addition of ethical components keeps the technical learning goals of the assignment the same. Asking instructors to consider when an algorithm is appropriate or what mistakes students might make if they apply an algorithm by rote can highlight potential impacts directly related to the instructor's technical learning goals. This reflection taught us an important lesson: identify technical nuances and student misconceptions, then choose an ethical context highlighting human judgments involved in these technical decisions. This process ultimately deepens the academic rigor associated with the learning goals, emphasizing nuances not previously made explicit. Once technical nuances are made explicit, it is easier to identify an ethical context. This context does not need to be complex or even highly realistic. It just needs to illustrate a human judgment in a right-and-wrong context where different technical decisions can have beneficial or harmful impacts. We discuss different ways to create such contexts in Sections 4.1.1, 4.1.2, and 4.1.3.

4.1.1 *Identify explicitly known human judgments involved in technical steps that could have potential impacts in a human context.*

There are many explicit technical decisions that a developer makes. In the school bus route optimization example, the explicit choice of optimization function affects the algorithm's training and results. Similarly, in the graph search algorithm example, the choice of algorithm involves an explicit human judgment that can impact the returned path. In a human context, path choice may have ethical ramifications. Contextualizing explicit technical decisions can reveal their potential impact on humans, making them a suitable focus for ethical integration.

4.1.2 *Identify implicit assumptions or rote calculations that could have downstream impacts in a human or real-world context if not carefully considered.*

Similarly, developers also make *unconscious* choices – implicit assumptions and rote calculations – with the potential to cause unforeseen impacts. These choices represent fine fissures in students' knowledge and are a promising target for potential ethical integration. Instructors will encounter and anticipate these fissures over time; unfounded assumptions and rote calculations may result in obvious errors [56]. Not only does the process of identifying these cracks enhance the learning goals, but it encourages the instructor to reflect on student misconceptions or assumptions, a technique that can help build an instructor's knowledge of how to teach the technical topic, also known as their Pedagogical Content Knowledge (PCK) [26, 35]. Instructional designers can work with course instructors to identify common rote calculations and assumptions, which may have unforeseen impacts if not carefully considered. For our example, the AI instructor identified students' rote use of A^* Search algorithms, which are guaranteed to find the path with the lowest cost in a graph assuming the chosen function to estimate the costs – the heuristic function – is admissible (the estimated cost from any node to the goal is never greater than the true cost of reaching the goal from that node). A rote use of A^* Search may return the lowest-cost path, but that path may not be the best option for a human in a given situation.

4.1.3 Look for existing real-world scenarios with potential ethical implications or human impacts. This was our initial approach, which hit a dead-end for many abstract topics, but can be effective for topics involved in newsworthy or personal experiences, such as the ones identified by Jarzemsky et al. [34]. These examples can highlight ethical decisions related to technical applications, applicable to classroom assignments in some cases. Moreover, real-world applications may improve student learning [5, 14, 53] and enable them to transfer their abstract knowledge to specific contexts [15]. The school bus example showed us that optimizing for one value (such as saving money) can ignore others (like equitable routes), leading to unforeseen consequences. This inspired us to consider a scenario where the standard algorithm may not be suitable. We contrasted a distance context with a safety context, where the least-safe segment determines the overall danger of a path. We contextualized our assignment by representing the nodes of the graph as cities and edges as roads, assigning the cost as a measure of safety. This problem illustrates how the lowest cumulative-cost path may be unsafe for a human. The full assignment is in Appendix B.

4.2 Challenge 2: Integrating Ethical Components while Maintaining a Technical Focus

Suggestion: Ensure that questions require both technical and ethical reasoning by examining the knowledge required for each task, and striving for increasing technical depth where students must change their technical answers based on their ethical decision making.

Maintaining a technical focus was difficult. In the first design iteration, we added a question about ethical considerations in our safety context: *Name one other factor you think would be ethically prudent to consider when designing an algorithm for a self-driving car.* Although it aimed to prompt students to consider that other factors affect route selection, such as the safety of the road, this question is only *related* to ethics in AI. That is, students can answer this question without drawing on or deepening AI content knowledge. To truly integrate professional ethics into the assignment, we realized we needed to create questions that demanded ethical considerations *and* technical knowledge.

However, integrating ethical components into technical questions took more work. For example, devising realistic heuristics for a safety context proved fruitless. While admissibility and consistency can be determined mathematically, heuristics based on Euclidean and Manhattan distances do not necessarily ensure safety. We realized we could require students to grapple with this; we asked them to explicitly consider heuristic appropriateness, questioning their suitability for the given context (Question 2, Appendix B). This breakthrough transformed the problem into a professional ethics question, contrasting with the original assignment that presented heuristics without context.

We learned a valuable lesson: examine the knowledge required for each task, and strive for increasing technical depth. To examine the face validity of our questions, we asked ourselves:

(1) Can this question be answered without using the AI knowledge that is targeted by this class?
(2) Does this question demonstrate the importance of human judgment in an ethics-relevant context?

The first question guided us away from our initial instinct to focus on the social harms that AI can cause, and the second prompted us to identify human judgment in algorithmic contexts. This last issue is subtle and lies at the heart of many ethical concerns in computing. To wit, it is often the case that specific and seemingly innocuous choices (or even default settings) made in the design of a system have broad ramifications when deployed in a social context, and it is only by making those choices visible that a designer can genuinely engage with the impacts those choices might have. The pedagogical key here is to help students recognize subtle technical choices.

To illustrate how we evaluate questions for ethics integration, we analyze our assignment questions (see Appendix B). Question 1 assesses technical proficiency by matching algorithms to paths. However, it does not demonstrate the importance of human judgment in an ethics-relevant context. Initially, we included this question to establish a baseline for the following ethics-integrated questions. Upon reflection, we have learned that integrating ethics into this question might highlight students' misconceptions. The instructor noted a significant mismatch in the safety context problem and traditional algorithms optimized for accumulation problems (e.g., A^* Search): if an individual segment is unsafe, the whole path is unsafe. That is, instead of minimizing the overall path cost, students had to consider minimizing the cost of the most expensive individual segment. The existing course assignments did not assess students' ability to recognize this difference, making it worthwhile to include. Though challenging, integrating ethics while maintaining a technical focus provided a more in-depth exploration of AI content.

Question 3 on our assignment requires students to use their AI knowledge of graph search algorithms to design a program that outputs the safest path by recommending algorithmic modifications to ensure passenger safety. This question successfully integrates ethics (ensuring safety) while maintaining a technical focus (suggesting algorithmic modifications) and highlights the importance of human judgment in this context. The think-aloud responses also support this claim, with participants demonstrating technical knowledge in response to the ethical context. P1 suggested:

> "I mean, if it was approaching [eight] or if it hit eight or more, just immediately stop and just go find a different path. Because why continue to find the total cost if it's already greater than eight when you just want to lead your passengers down a safe path?"

In this example, the participant adjusted their algorithmic decision to avoid unsafe paths with segment costs of eight or more (the question noted that costs of eight indicated extremely unsafe roads). The question required the student to draw on their technical knowledge of how graph search algorithms procedurally reason through a path while challenging the student to adjust their decisions based on the context. Similarly, P2 responded:

> "What would you do to ensure your program does not lead passengers down an unsafe path? Well, I guess the easiest thing is to submit the path for review modification. Seems like the easiest thing to do is to just have a human look at it and say, 'Hey, I don't want to drive down this road because I'm going to die.' So, I'll have the person do that. I guess it's also possible [for] any path greater than eight, modify the cost of it. I can also just make it prohibitively expensive to go down anything greater than eight."

P2's response shows they used technical knowledge (path cost determination) and ethical reasoning (human-in-the-loop approach). Such questions allow students to practice making technical decisions in response to ethical concerns. By integrating ethics while maintaining a technical focus, our assignments highlighted professional ethics concerning the CS content the students are learning in class.

We propose that a good ethics-integrated question should leave room for technically plausible but contextually illogical answers, requiring students to change their technical decisions in response to the human context. There may be cases where students do not need to adjust their technical decisions to respond to an ethical component, even when a problem is contextualized (e.g., if a standard implementation of the A^* Search algorithm returned the safest path for a graph). In these cases, whether students consciously used their chosen technical approach or simply followed a standard procedure would be unclear. In contrast, a human context that requires a different approach than an abstract context can make students' decision-making process explicit.

4.3 Challenge 3: Helping Students Integrate Technical and Ethical Thinking

> **Suggestion**: Anticipate students' tendency to answer questions from either a purely technical or purely philosophical perspective, and write clear instructions that explicitly encourage them to use both technical and ethical modes of reasoning to answer each question.

Think-alouds with participants revealed challenges in prompting students to use both technical and ethical thinking. Understanding the tasks proved challenging due to the mix of technical and non-technical terms, particularly in cases where technical terms had non-technical meanings. For example, P2 described *"reasonable"* as *"something a little more arbitrary"* than 'admissibility' and 'consistency' during the think-aloud and, in a debriefing session after completing the assignment, asked:

> "Am I am right in thinking that a 'reasonable' heuristic is something I need to argue, as opposed to having a concrete definition?"

P3 also struggled with this idea while completing the assignment:

> "I guess I should probably assume, or I'm working under the assumption that I should just go with this idea that we're talking about 'admissible' in the context of just straightforward admissibility of heuristics in AI, not about the ethical admissibility of things. But I would probably end up giving very different answers if we're talking about the ethical admissibility."

They explained this thought in more detail during the debriefing session:

> "For me, whenever I see a keyword like 'admissible' or 'consistent' and I know that it has a definition in the formal study of graph theory, I automatically think, 'Okay, so I'm going into STEM mode,' I'm not thinking about it as an ethics problem anymore. Whereas when we talk about the danger of roads, [...] it feels like it could be an ethics question - here's this buzzword or here's this keyword. [...] It seems like I have these two conflicting ideas of what that question could be about. But it wasn't clear to me one way or the other."

Since many CS courses teach students to solve problems in "STEM mode," students may not be prepared to integrate ethical considerations into their technical problem-solving approach. Further, as instructional designers, we may struggle to write questions that elicit both "STEM" and ethical reasoning or encourage seamless switching between the two modes. For instance, Question 2b (see Appendix B) aimed to test students' ability to calculate returned paths and assess their reasonableness based on the context. Using the lessons we learned in Challenge 2 (Section 4.2), we thought this question would require students to draw on both modes of reasoning. However, P3's response revealed a communication gap in our intended approach:

> "The thought just occurred to me - if you have a proper implementation of any of these algorithms, it should still get you the best one. It's just a matter of how quickly it gets you there. So I feel like under that criteria, or approaching it that way, every single one of these should be yes [that all of the listed paths are reasonable for a human driver to take]."

The participant in this example solely relied on technical knowledge and overlooked the safety context, assuming that a proper implementation guarantees a reasonable path without considering the consequences of the technical decision. This concept underscores our primary objective of integrated assignments: enabling students to scrutinize technical decisions that may otherwise be unquestioned. Mere computations can result in harmful suggestions, even with an appropriately optimized algorithm, such as proposing that a driver take an unsafe path. Interestingly, while this participant interpreted this question in "STEM mode," P2 answered this same question through a purely ethical lens:

> "[Which path(s) would be reasonable for a human to traverse?] I guess that greatly depends on what the person is doing. If you are going to visit family, you might as well stay home. If you're delivering vaccines to a rural village, that might be something different."

This participant's response focused on the philosophical ethics aspect of the question, but we intended for it to require both technical feasibility and appropriateness for the human context. We aim to encourage dual-mode thinking in our assignments, but we recognize that different students may interpret the questions differently. This lesson was reinforced when evaluating student responses to the instructor-modified AI course assignment (Appendix C). Despite attempts to clarify the instructions, many students answered the questions without technical reasoning, treating them purely as philosophical questions [8]. Since the assignment had a human context and was distinct from their typical abstract assignments, some students may not have realized that technical knowledge was necessary. In the future, we can improve our questions to evoke the intended dual-mode thinking, for instance, by separating technical terms from ethical components (e.g., clarifying that 'reasonable' refers to a decision made by a human). Instructors can also introduce the assignment by noting that the questions require both technical and ethical reasoning. Future work can explore additional design guidelines to create ethics-integrated questions that consistently elicit answers drawing on both modes of thinking.

4.4 Challenge 4: Making the Assignment Practical for the Classroom

> **Suggestions**: (1) Scale the assignment using auto-gradable questions, and (2) prepare for errors in students' technical execution that could result in incorrect or harmful reasoning about the broader ethical issues.

4.4.1 Scaling the assignment for assessment in large classes. Designing ethically-integrated assignments is challenging; scaling them for large classes can be even harder. Our initial draft had many open-response questions, which are time-consuming to grade and would have been cumbersome for the course staff grading over 120 assignments. We revised the draft to include multiple-choice and numerical questions while retaining some open-response items. Our think-alouds revealed the value of these formats for prompting and revealing ethics-relevant thinking in AI without limiting the range of student responses, consistent with prior studies [69]. In many cases, including all possible answer choices as multiple-choice options is possible.

As an illustration, our second assignment contextualized a Bayesian probability problem using youth drug arrests based on race (Appendix D). On this assignment, students were tasked with computing the ratio of conditional probabilities, revealing that African American youth are arrested roughly 80% more frequently than white youth, in proportion to their population share. Question 7 asks:

> Which of the following is true, based on your calculations?
> (a) White youth are being arrested disproportionately more often than African American youth, relative to their share of the population.
> (b) African American youth are being arrested disproportionately more often than white youth, relative to their share of the population.
> (d) African American youth and white youth are being arrested at proportionally the same rate, relative to their share of the population.
> (d) There is not enough information to tell if any of the above are true.

The multiple-choice options for this question provide a comprehensive range of potential answers. Although the multiple-choice format did not constrain students' answer choices, our think-alouds demonstrated that different students might select the same option for vastly different reasons. In this instance, P1 and P2 opted for answer (d) *there is not enough information to tell if any of the above are true.* However, their rationales for choosing this response were distinct. P1 justified their selection as follows:

> "Given that those are just probabilities for the measurements of African American and white people, and also the fact that earlier it said that there weren't any statistics on youth explicitly, I don't think that there's enough information to tell if any of the three above are true, because you would just be making assumptions, and that isn't necessarily a good thing in this situation, especially if you're trying to explicitly target population shares with specifically youth."

This participant reasoned that since the statistics provided were for the entire population and not specific to youth, the question was unanswerable. Although P2 agreed with this answer choice, their rationale differed. They argued that the computations failed to consider population shares entirely:

> "This one I know for sure there's not enough information, because we're not talking at all about population ratios."

For future iterations, we plan to add multiple-choice questions that prompt students to justify their conclusions. We can achieve this by having one group of students answer an open-response version of the question and then use the most frequent responses to create multiple-choice options. This method can indicate if students are arriving at correct answers via informed and accurate computations rather than random guessing, while maintaining an auto-gradable format.

4.4.2 Potential discomfort when using a real-world context. Another crucial aspect is the need for a comfortable and respectful learning environment that avoids trivializing serious real-world issues. On a subsequent question in the Bayesian probability assignment (Question 8, Appendix D), students are presented with incidence rates for drug-related crimes, revealing that African American youth exhibit lower drug use and sales rates than white youth. However, during the think-aloud with P2, the interviewer (South), a Person of

Color, felt uncomfortable when P2 arrived at the incorrect conclusion that white youth, rather than African American youth, were subject to disproportionate arrests. While we as a design team acknowledged the potential discomfort arising from discussions of systemic racism, we did not anticipate the mathematical error P2 made and, therefore, their uncomfortable conclusion. After extensive discussions, we concluded that understanding systemic racism outweighed the potential for discomfort and that a real-world example was necessary to demonstrate the consequences of accurate and inaccurate calculations. Although this study does not focus on assessing the discomfort associated with such assignments, we suggest that future work analyze their emotional impacts. If the context involves sensitive and serious topics, like systemic racism, instructors must be prepared for technical execution errors that may result in incorrect and harmful reasoning about broader ethical issues. To address this, we recommend following assignments with in-class discussions to prevent important topics or emotions from being overlooked. Jarzemsky et al. [34] proposed a potential solution using hypothetical situations on assignments, where students can discuss how the problem might apply to real-world controversies in class (e.g., an assignment about removing dog content from a social media platform for cats may lead to discussions about real-world censorship, filtering bias, and misinformation [34]). To ensure a comfortable and respectful learning environment while addressing ethical controversies, we encourage future research to explore additional strategies beyond those discussed in this study.

5 DISCUSSION

The challenges and opportunities revealed by the RtD method are specific to our conception of ethics. Across the CS education community, ethics is conceptualized in myriad ways, ranging from discussions of ACM's code of ethics [24], to philosophical frameworks (e.g., [16, 32, 58]), to technical considerations regarding privacy [62], security [20], and accessibility [65], to overarching issues of social justice [45], and beyond [64]. Our conception of ethics, professional ethics (discussed in Section 2.1) [36], focuses on the responsibility of the CS expert to consider if a technically-feasible implementation is matched to the particular problem context. The challenges identified by our process echo well-known obstacles to teaching ethics in CS, such as the difficulty of connecting ethics to abstract or mathematical content [64], while presenting details that have not been discussed in prior work. We expect that these high-level challenges will apply across the range of ethics conceptions, but will manifest differently, illuminating different aspects of these challenges and pointing to different solutions for different contexts. In the next sections, we discuss how professional ethics can address challenges in connecting ethics to abstract content and in ensuring sufficient class time for technical topics (Section 5.1). Next, we discuss how that framing can be combined with other conceptions of ethics (Section 5.2). Finally, we argue for the benefits of RtD in supporting the integration of ethics in CS (Section 5.3).

5.1 Ethics Can Be Synergistic with Technical Topics

Our first challenge, the difficulty of finding ethical contexts for arbitrary technical topics (described in Section 4.1), is one that many instructors have faced [31, 64]. However, this challenge can be addressed by shifting our instructional design process: instead of matching real-world ethical examples to course content, start by identifying human judgments involved in the technical topic (e.g., assumptions, algorithmic starting conditions), then choose a context that highlights the potential outcomes of these judgments. This reframing may allow us to connect ethical concepts to significantly more technical topics, including abstract or mathematical ones, without changing the course content. Making connections between ethics and technology through the judgements inherent in technical implementation also addresses other perceived barriers: limited time, competing priorities, and lack of control over the curriculum [31, 64]. These challenges suggest that instructors see technical and ethical topics as vying for class time and focus. However, integrating ethics need not detract from technical topics and may even enhance technical rigor. In Challenge 2 (Section 4.2), we discussed how questions that focus on technical material while emphasizing the importance of human judgment effectively integrated ethics while highlighting previously unassessed technical misconceptions. With this approach, instructors can incorporate professional ethics in a small way without needing additional time or a curriculum overhaul.

5.2 Improving Technical and Ethical Integration

Challenge 3 (Section 4.3) revealed that students struggled to integrate technical and ethical thinking when solving tech-ethics problems. While students may improve this skill with repeated exposure to such assignments, exploring and refining question design can help elicit the intended dual-mode thinking. Further, assignments designed around professional ethics may provide useful links between technical skills and broader ethical issues, helping students understand the connections between technical choices and socio-technical outcomes, and also helping students recognize what types of tech-ethics issues should (or cannot) be addressed with technical knowledge. To encourage this integration, we echo the call from Goetze [27] for transdisciplinary collaboration between computing and philosophy experts to design assignments that require simultaneous reliance on both modes of reasoning. Acknowledging that our team consisted only of experts in computing, future RtD work can explore how collaboration with philosophy experts improves the design process.

In considering how to support students in connecting professional-ethics choices to broader ethical and socio-technical issues, we note that these broader issues are situated in particular cultural structures and power dynamics. Technical assignments alone will not teach students to understand the social and cultural impact of computing in response to ever-evolving societal forces or political structures [42, 50]. For example, a student can successfully solve our Bayesian probability assignment (Appendix D) without considering how their solutions respond to the broader social context of systemic racism and historical inequities. To support justice-centered computing efforts [45, 67], we urge instructors to supplement assignments with additional information on how technical choices shape society, acknowledging current social structures and historical context, along with recognizing and challenging our own biases, attitudes, and beliefs [70].

5.3 The Promise of Research through Design for Ethics in CS

Instructional design is the link between the high-level goal of integrating ethics into computing and the actual delivery of instruction and assessment in the classroom. While some ethics-integrated assignments can be reused across courses, many CS topics do not yet have examples of ethics-integrated material [64] — material which can only be generated through instructional design. Therefore, we argue that sharing and studying design processes is as important as sharing and studying specific instructional materials. Further, we expect that shared instructional material will be made more valuable when accompanied by the design considerations that shaped them. Being explicit about design choices may make it easier for others to adapt the materials to their own specific needs. Finally, when researchers propose ways to incorporate ethics into CS (as we do above), we recommend advocating for research on the design process itself. Understanding instructors' challenges in following different methods for integrating ethics in CS is a necessary step for creating design processes that instructors can use.

6 LIMITATIONS AND FUTURE WORK

Studying our own experiences involves inherent limitations. Although we drew from contemporaneous written notes in analyzing our experiences, our accounts of our own experiences are inherently subjective, and both our initial actions and our reflections on them are shaped by our individual biases. While we anticipate that our challenges and lessons learned will be applicable across topics in the computing curriculum, further work is necessary to empirically validate this claim. Additionally, our lessons learned are suggestions — we did not evaluate their usefulness in mitigating any of the challenges, explore if others could follow them, or examine if they result in high-quality materials. And, as noted in the Discussion (Section 5), our identified challenges and lessons learned are likely specific to our choice of professional ethics as a frame. Many of these limitations point to avenues for future work.

One of the facets of the RtD method is repetition (Step 5, described in Section 2.3) [75]. We therefore view this research as ongoing, and the design suggestions we presented in our results can be expanded using the results of additional design studies. We encourage future work to consider design research so we can uncover additional reflections and learn from the design process itself.

7 CONCLUSION

In this work, we describe practical suggestions to overcome challenges involved in creating ethically-integrated assignments for computing courses: (1) select an ethical context that emphasizes human judgments involved in technical nuances, (2) craft questions that require both ethical and technical reasoning, (3) explicitly prompt students to reason from both ethical and technical perspectives, and (4) make the assignment feasible for the classroom by using auto-gradable questions and accounting for potential errors in student reasoning. These insights stem from our experience developing and testing such assignments in an AI course and can aid in designing new ethically-integrated assignments or integrating ethical concepts into existing technical assignments. We hope these lessons make it easier for computing instructors to incorporate ethical concepts within their courses without compromising technical content.

ACKNOWLEDGMENTS

We thank Mu Zhang and our partner instructor, Alan Kuntz, for their valuable contributions to the design of our assignments. We also thank Lauren Margulieux and members of the HCC writing seminar at the University of Utah for their thoughtful feedback on drafts of this work. Additionally, we used ChatGPT to shorten our submission (by inputting our original text from Sections 3 and 4 and asking for shortened versions). This work was supported by the National Science Foundation (NSF award SaTC-EDU 2041960), and by the Mozilla Foundation through a Responsible Computing Challenge award to the University of Utah.

REFERENCES

[1] Substance Abuse and Mental Health Services Administration. 2005. Results from the 2004 National Survey on Drug Use and Health: National Findings. *Office of Applied Studies, NSDUH Series H-28, DHHS Publication No. SMA 05-4062* (2005).
[2] Ashish Aggarwal and Saurabh Ranjan. 2022. How do Undergraduate Students Reason about Ethical and Algorithmic Decision-Making?. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1*. 488–494.
[3] Robert M Aiken. 1983. Reflections on teaching computer ethics. *ACM SIGCSE Bulletin* 15, 3 (1983), 8–12.
[4] Davey Alba. 2022. CHATGPT, open ai's chatbot, is spitting out biased, sexist results. https://www.bloomberg.com/news/newsletters/2022-12-08/chatgpt-open-ai-s-chatbot-is-spitting-out-biased-sexist-results
[5] Steve Olusegun Bada and Steve Olusegun. 2015. Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education* 5, 6 (2015), 66–70.
[6] Dimitris Bertsimas, Arthur Delarue, and Sebastien Martin. 2019. From school buses to start times: Driving policy with optimization. In *Informs Business Analytics Conference, April*. 14–16.
[7] Neal E. Boudette. 2021. Tesla's Autopilot Technology Faces Fresh Scrutiny. https://www.nytimes.com/2021/03/23/business/teslas-autopilot-safety-investigations.html
[8] Noelle Brown, Koriann South, and Eliane S Wiese. 2022. The Shortest Path to Ethics in AI: An Integrated Assignment Where Human Concerns Guide Technical Decisions. In *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1*. 344–355.
[9] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.
[10] Emanuelle Burton, Judy Goldsmith, Sven Koenig, Benjamin Kuipers, Nicholas Mattei, and Toby Walsh. 2017. Ethical considerations in artificial intelligence courses. *AI Mag.* 38, 2 (July 2017), 22–34.
[11] Emanuelle Burton, Judy Goldsmith, and Nicholas Mattei. 2016. Using "the machine stops" for teaching ethics in artificial intelligence and computer science. *Association for the Advancement of ArtificialIntelligence* (2016).
[12] John M Carroll and Wendy A Kellogg. 1989. Artifact as theory-nexus: Hermeneutics meets theory-based design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 7–14.
[13] Sharon Carver. 2001. Cognition and Instruction: Enriching the Laboratory School Experience of Children, Teachers, Parents, and Undergraduates. In *Cognition and instruction: Twenty-five years of progress*, S M Carver and D Klahr (Eds.). Lawrence Erlbaum Associates, Inc., Mahwah, NJ.
[14] Kevin R Clark. 2018. Learning theories: constructivism. *Radiologic Technology* 90, 2 (2018), 180–182.
[15] National Research Council. 2000. *Learning and Transfer: Chapters 2 and 3 How People Learn: Brain, Mind, Experience, and School: Expanded Edition.* The National Academies Press, Washington, DC.
[16] Bill Cowan, Elodie Fourquet, and Marta Kryven. 2018. Teaching the societal consequences of computer science: new ideas for increasing student involvement. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*. 242–247.
[17] Michael Davis. 2006. Integrating ethics into technical courses: micro-insertion. *Sci. Eng. Ethics* 12, 4 (Oct. 2006), 717–730.
[18] Stacy A Doore, Casey Fiesler, Michael S Kirkpatrick, Evan Peck, and Mehran Sahami. 2020. Assignments that Blend Ethics and Technology. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (Portland, OR,

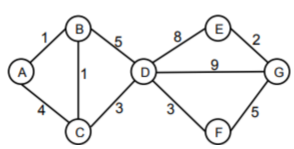
USA) *(SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 475–476.
[19] Minette Drumwright, Robert Prentice, and Cara Biasucci. 2015. Behavioral ethics and teaching ethical decision making. *Decis. Sci. J. Innov. Educ.* 13, 3 (July 2015), 431–458.
[20] Richard G Epstein. 2006. An ethics and security course for students in computer science and information technology. *ACM SIGCSE Bulletin* 38, 1 (2006), 535–537.
[21] Casey Fiesler, Mikhaila Friske, Natalie Garrett, Felix Muzny, Jessie J Smith, and Jason Zietz. 2021. Integrating Ethics into Introductory Programming Classes. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education.* Association for Computing Machinery, New York, NY, USA, 1027–1033.
[22] Casey Fiesler, Natalie Garrett, and Nathan Beard. 2020. What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (Portland, OR, USA) *(SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 289–295.
[23] Kathi Fisler, Sorelle Friedler, Kevin Lin, and Suresh Venkatasubramanian. 2022. Approaches for Weaving Responsible Computing into Data Structures and Algorithms Courses. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2.* 1049–1050.
[24] Association for Computing Machinery. 2018. ACM Code of Ethics and Professional Conduct. https://www.acm.org/code-of-ethics
[25] Natalie Garrett, Nathan Beard, and Casey Fiesler. 2020. More Than "If Time Allows": The Role of Ethics in AI Education. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* Association for Computing Machinery, New York, NY, USA, 272–278.
[26] Julie Gess-Newsome. 1999. Pedagogical content knowledge: An introduction and orientation. *Examining pedagogical content knowledge: The construct and its implications for science education* (1999), 3–17.
[27] 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* (Toronto ON, Canada) *(SIGCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 645–651.
[28] Judy Goldsmith and Emanuelle Burton. 2017. Why teaching ethics to AI practitioners is important. In *Thirty-First AAAI Conference on Artificial Intelligence.*
[29] Ellen P Goodman. 2019. The challenge of equitable algorithmic change. *The Regulatory Review, Feb* (2019).
[30] Colin M Gray, Shruthi Sai Chivukula, Kassandra Melkey, and Rhea Manocha. 2021. Understanding "Dark" Design Roles in Computing Education. In *Proceedings of the 17th ACM Conference on International Computing Education Research* (Virtual Event, USA) *(ICER 2021)*. Association for Computing Machinery, New York, NY, USA, 225–238.
[31] Colleen Greer and Marty J Wolf. 2020. Overcoming barriers to including ethics and social responsibility in computing courses. In *Societal Challenges in the Smart Society.* Universidad de La Rioja, 131–144.
[32] Amir Hedayati-Mehdiabadi. 2022. How do computer science students make decisions in ethical situations? Implications for teaching computing ethics based on a grounded theory study. *ACM Transactions on Computing Education (TOCE)* 22, 3 (2022), 1–24.
[33] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. Most people are not WEIRD. *Nature* 466, 7302 (July 2010), 29. https://doi.org/10.1038/466029a
[34] Julie Jarzemsky, Joshua Paup, and Casey Fiesler. 2023. "This Applies to the Real World": Student Perspectives on Integrating Ethics into a Computer Science Assignment. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (Toronto ON, Canada) *(SIGCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 374–380.
[35] HU Jing-Jing. 2014. A critical review of pedagogical content knowledge'components: Nature, principle and trend. *International Journal of Education and Research* 2, 4 (2014), 411–424.
[36] Deborah G Johnson. 1994. *Computer ethics (2nd ed.).* Prentice-Hall, Inc., USA.
[37] LD Johnston, PM O'Malley, JG Bachman, and JE Schulenberg. 2005. Monitoring the Future Occasional Paper No. 61. *Ann Arbor, MI: Institute for Social Research* (2005), 1975–2004.
[38] Van Jones. 2017. Are blacks a criminal race? surprising statistics. https://www.huffpost.com/entry/are-blacks-a-criminal-rac_b_8398
[39] NC Keramaris, NK Kanakaris, Chris Tzioupis, G Kontakis, and PV Giannoudis. 2008. Translational research: from benchside to bedside. *Injury* 39, 6 (2008), 643–650.
[40] Shamika Klassen and Casey Fiesler. 2022. "Run Wild a Little With Your Imagination": Ethical Speculation in Computing Education with Black Mirror. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1* (Providence, RI, USA) *(SIGCSE 2022)*. Association for Computing Machinery, New York, NY, USA, 836–842.
[41] Amy J Ko. 2019. 21st Century Grand Challenges for Computing Education. 19th Koli Calling Conference on Computing Education Research.
[42] Amy J Ko, Alannah Oleson, Neil Ryan, Yim Register, Benjamin Xie, Mina Tari, Matthew Davidson, Stefania Druga, and Dastyni Loksa. 2020. It is time for more critical CS education. *Commun. ACM* 63, 11 (Oct. 2020), 31–33.
[43] Kenneth R Koedinger, Albert T Corbett, and Charles Perfetti. 2012. The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science* 36, 5 (2012), 757–798.
[44] Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin. 2016. How we analyzed the COMPAS recidivism algorithm. *ProPublica (5 2016)* 9, 1 (2016), 3–3.
[45] Kevin Lin. 2022. CS Education for the Socially-Just Worlds We Need: The Case for Justice-Centered Approaches to CS in Higher Education. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1* (Providence, RI, USA) *(SIGCSE 2022)*. Association for Computing Machinery, New York, NY, USA, 265–271.
[46] Marcia C Linn, H S. Lee, R Tinker, F Husic, and J L Chiu. 2006. Teaching and Assessing Knowledge Integration in Science. *Science* 313 (2006), 1049–1050. https://doi.org/10.1126/science.1131408
[47] K M MacQueen, E McLellan-Lemal, K Bartholow, and B Milstein. 2008. Team-based codebook development: Structure, process, and agreement. *Cultural Anthropology Methods* 10, 2 (Jan. 2008), 119–135.
[48] James W Malazita and Korryn Resetar. 2019. Infrastructures of abstraction: how computer science education produces anti-political subjects. *Digital Creativity* 30, 4 (Oct. 2019), 300–312.
[49] Pru Mitchell. 2016. From concept to classroom: What is translational research? (2016).
[50] Jared Moore. 2020. Towards a more representative politics in the ethics of computer science. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.* 414–424.
[51] Alex Najibi. 2020. Racial Discrimination in Face Recognition Technology. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/
[52] B Nalini. 2019. The Hitchhiker's Guide to AI Ethics Part 3: What AI Does & Its Impact. https://towardsdatascience.com/the-hitchhikers-guide-to-ai-ethics-part-3-what-ai-does-its-impact-c27b9106427a
[53] Hilarie Nickerson, Catharine Brand, and Alexander Repenning. 2015. Grounding Computational Thinking Skill Acquisition Through Contextualized Instruction. In *Proceedings of the eleventh annual International Conference on International Computing Education Research* (Omaha, Nebraska, USA) *(ICER '15)*. Association for Computing Machinery, New York, NY, USA, 207–216.
[54] National Household Survey on Drug Abuse. 1999. The Substance Abuse and Mental Health Services Administration, Office of Applied Studies, Table G. *Office of Applied Studies, NSDUH Series H-28, DHHS Publication No. SMA 05-4062* (1999).
[55] Evan Peck. 2018. *Ethical Reflection Modules for CS 1.* Technical Report. Retrieved 2018-10-13 from https://ethicalcs.github.io.
[56] Yizhou Qian and James Lehman. 2017. Students' Misconceptions and Other Difficulties in Introductory Programming: A Literature Review. *ACM Trans. Comput. Educ.* 18, 1 (Oct. 2017), 1–24.
[57] Gordon Rowland. 1993. Designing and instructional design. *Educational technology research and development* 41, 1 (1993), 79–91.
[58] Jeffrey Saltz, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. 2019. Integrating Ethics within Machine Learning Courses. *ACM Trans. Comput. Educ.* 19, 4 (Aug. 2019), 1–26.
[59] Jeffrey S Saltz, Neil I Dewar, and Robert Heckman. 2018. Key Concepts for a Data Science Ethics Curriculum. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) *(SIGCSE '18)*. Association for Computing Machinery, New York, NY, USA, 952–957.
[60] Kathryn Elizabeth Shroyer and Jennifer A Turns. 2021. Research Through Design: A Promising Methodology for Engineering Education. In *2021 ASEE Virtual Annual Conference Content Access.*
[61] Michael Skirpan, Nathan Beard, Srinjita Bhaduri, Casey Fiesler, and Tom Yeh. 2018. Ethics Education in Context: A Case Study of Novel Ethics Activities for the CS Classroom. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) *(SIGCSE '18)*. Association for Computing Machinery, New York, NY, USA, 940–945.
[62] Michael Skirpan, Jacqueline Cameron, and Tom Yeh. 2018. Quantified self: An interdisciplinary immersive theater project supporting a collaborative learning environment for cs ethics. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education.* 946–951.
[63] Mona Sloane. 2019. Inequality Is the Name of the Game: Thoughts on the Emerging Field of Technology, Ethics and Social Justice. In *Weizenbaum Conference.* DEU, 9.
[64] Jessie J Smith, Blakeley H Payne, Shamika Klassen, Dylan Thomas Doyle, and Casey Fiesler. 2023. Incorporating Ethics in Computing Courses: Barriers, Support, and Perspectives from Educators. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (Toronto ON, Canada) *(SIGCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 367–373.
[65] Chia-En Tseng, Seoung Ho Jung, Yasmine N Elglaly, Yudong Liu, and Stephanie Ludi. 2022. Exploration on Integrating Accessibility into an AI Course. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1* (Providence, RI, USA) *(SIGCSE 2022)*. Association for Computing Machinery, New York, NY, USA, 864–870.

[66] UC Berkeley. 2014. UC Berkeley CS188 Intro to AI – Course Materials. http://ai.berkeley.edu/home.html, Accessed September 2020.
[67] Sepehr Vakil. 2018. Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education. *Harvard Educational Review* 88, 1 (2018), 26–52.
[68] Pranshu Verma. 2022. These robots were trained on AI. they became racist and sexist. https://www.washingtonpost.com/technology/2022/07/16/racist-robots-ai/
[69] Xu Wang, Carolyn Rose, and Ken Koedinger. 2021. Seeing Beyond Expert Blind Spots: Online Learning Design for Scale and Quality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21, Article 51)*. Association for Computing Machinery, New York, NY, USA, 1–14.
[70] Alicia Nicki Washington. 2020. When Twice as Good Isn't Enough: The Case for Cultural Competence in Computing. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (Portland, OR, USA) *(SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 213–219.
[71] Jordan Weissmann. 2018. Amazon created a hiring tool using AI. it immediately started discriminating against women. https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html
[72] Grant Wiggins and Jay McTighe. 2005. Gaining Clarity on Our Goals. In *Understanding by Design*, Association for Supervision & Curriculum Development (Ed.). ProQuest Ebook Central.
[73] Ellen D Wu. 2014. *The color of success: Asian Americans and the origins of the model minority.* Princeton University Press.
[74] Cass Zegura, Ben Rydal Shapiro, Robert MacDonald, Jason Borenstein, and Ellen Zegura. 2023. "Moment to Moment": A Situated View of Teaching Ethics from the Perspective of Computing Ethics Teaching Assistants. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23-28, 2023, Hamburg, Germany* 1, 1 (2023).
[75] John Zimmerman and Jodi Forlizzi. 2014. Research through design in HCI. In *Ways of Knowing in HCI*. Springer, 167–189.

A ORIGINAL GRAPH SEARCH ASSIGNMENT (PRIOR TO ETHICS-INTEGRATION)

This version of the assignment is from the Introduction to AI course at the University of California, Berkeley [66].

Consider the state space graph shown below. A is the start state and G is the goal state. The costs for each edge are shown on the graph. Each edge can be traversed in both directions. Note that the heuristic h_1 is consistent but the heuristic h_2 is not consistent.



Node	h_1	h_2
A	9.5	10
B	9	12
C	8	10
D	7	8
E	1.5	1.5
F	4	4.5
G	0	0

(1) Possible paths returned

For each of the following graph search strategies (*do not answer for tree search*), mark which, if any, of the listed paths it could return. Note that for some search strategies the specific path returned might depend on tie-breaking behavior. In any such cases, make sure to mark *all* paths that could be returned under some tie-breaking scheme.

Search Algorithm	Paths
Depth first search	A-B-D-G
Breadth first search	A-C-D-G
Uniform cost search	A-B-C-D-F-G
A^* search with heuristic h_1	
A^* search with heuristic h_2	

(2) Heuristic function properties

Suppose you are completing the new heuristic function h_3 as shown. All the values are fixed except $h_3(B)$. For each of the following conditions, write the set of values that are possible for $h_3(B)$. For example, to denote all non-negative numbers, write $[0, \infty]$, to denote the empty set, write $\emptyset$, and so on.

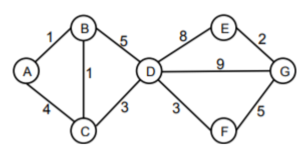
Node	A	B	C	D	E	F	G
h_3	10	?	9	7	1.5	4.5	0

(a) What values of $h_3(B)$ make h_3 admissible?
(b) What values of $h_3(B)$ make h_3 consistent?
(c) What values of $h_3(B)$ will cause A^* graph search to expand node A, then node C, then node B, then node D in order?

B ETHICS-INTEGRATED GRAPH SEARCH ASSIGNMENT

Consider the state space graph below. A is the starting city and G is the city where you want to end up. The edges represent roads between cities. The costs for each edge are shown on the graph. Each edge can be traversed in both directions. The A^* search algorithm will use the following heuristics, shown in the table below:

- h_1 represents the straight-line distance between the given location and G.
- h_2 represents the distance calculated as latitudinal distance + longitudinal distance from the given location to location G.



Node	h_1	h_2
A	9.5	10
B	9	12
C	8	10
D	7	8
E	1.5	1.5
F	4	4.5
G	0	0

(1) The costs represent travel time in hours between two cities. Match each algorithm to all paths it could return (if any)

Search Algorithm	Paths
Depth first search	A-B-D-G
Breadth first search	A-C-D-G
Uniform cost search	A-B-C-D-F-G
A^* search with heuristic h_1	
A^* search with heuristic h_2	

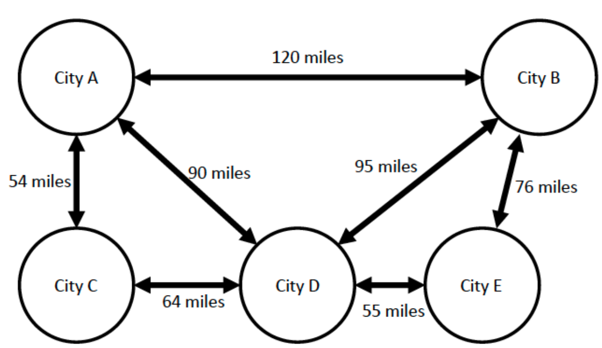
(2) Now let us consider the case where the costs represent passenger safety along the path between two cities. A cost of 1 means the path is as safe as possible, while a cost of 8 or more means it is very likely the passenger will die if they take that path.
  (a) Which of the following remain true?
    - h_1 is an admissible heuristic
    - h_1 is a consistent heuristic
    - h_1 is a reasonable heuristic to use for this problem
    - h_2 is an admissible heuristic
    - h_2 is a consistent heuristic
    - h_2 is a reasonable heuristic to use for this problem
  (b) For each of the following paths, mark whether or not it is a reasonable path for a human driver to take (Y/N): A-B-D-G, A-C-D-G, A-B-C-D-F-G.

(3) Consider the problem from (2) again.
  (a) Which one of these algorithms could you use to design a program that will take as input a new map, and output the safest path from a given start node to a given end node?
    - Depth First Search
    - Breadth First Search
    - Uniform Cost Search
    - A^* search with heuristic h_1, calculated to match h_1's description. (i.e. Straight line distance.)
    - A^* search with heuristic h_2, calculated to match h_2's description. (i.e. Latitudinal + Longitudinal distance.)
  (b) What would you do to ensure your program does not lead passengers down an unsafe path? (i.e. a path containing at least one edge with a cost of 8 or more.)

C ETHICS-INTEGRATED GRAPH SEARCH ASSIGNMENT FOR THE CLASSROOM

More details about this version of the assignment, including an analysis of student responses, can be found in our prior work [8].

For this section, we are asking for thoughtful consideration of the questions. Different answers can be correct, as long as they consider the algorithms in light of human-centric factors. Consider the following graph, which shows the connectivity via roads between five cities. The edges and their labels represent the existence and length of the roadways in between the cities.



Jessie, who resides in City A, needs to travel to City B. Jessie has a medical condition that frequently produces life-threatening symptoms at unpredictable times that requires treatment in an emergency room. There are emergency rooms in each of the five cities, but medical care is non-existent on the roads themselves. Please consider and thoughtfully answer the following questions:

(1) Is shortest path the right cost function when considering the optimal route for Jessie from City A to City B? Why or why not?
(2) Please describe an alternative cost function that may be appropriate for this problem and justify the choice.
(3) What is an optimal path from A to B in the above graph under your alternative cost function?
(4) Which, if any, of the search algorithms we've covered will still be optimal for your cost function, both for this specific graph and for an arbitrary graph? Justify your answer.
(5) What is another example of a human-centric factor that demonstrates the complex nature of defining an "optimal" path in real life?

D ETHICS-INTEGRATED BAYESIAN PROBABILITY ASSIGNMENT

You are working for the criminal justice system and one day your boss gives you the following probabilities:

- The probability that an African American youth aged 12-17 has been arrested for a drug offense, $P(D|A) = 0.00314$.
- The probability that a white youth aged 12-17 has been arrested for a drug offense, $P(D|W) = 0.00175$.
- The probability that a randomly selected person in America is African American, $P(A) = 0.123$.
- The probability that a randomly selected person in America is white, $P(W) = 0.751$.

Your boss says that since P(A,D) is less than P(W,D), this means fewer African American youth are being arrested than white youth, and this means they are not being arrested disproportionately. The following questions will help you determine whether this is true or not:

(1) Find $P(W, D)$, the probability that a randomly selected person in America is a white youth aged 12-17 and they have been arrested for a drug offense at least once. We don't have the probabilities of a person being a white/African American youth in America, but we do have the probabilities of a person being a white/African American person in America so just use those.
(2) Find $P(A, D)$, the probability that a randomly selected person in America is an African American youth aged 12-17 and they have been arrested for a drug offense at least once. We don't have the probabilities of a person being a white/African American youth in America but we do have the probabilities of a person being a white/African American person in America; use those.
(3) Was your boss' math correct? Yes / No.
(4) How does $P(D|A)$ compare to $P(D|W)$? To answer this, find x in $P(D|A) = xP(D|W)$, and then explain what x means.
(5) If your boss' math was correct, was their claim correct based on the comparison of $P(A, D)$ and $P(W, D)$?
(6) Which of the following is true, based on your calculations?

(a) White youth are being arrested disproportionately more often than African American youth, relative to their incidence of crime
(b) African American youth are being arrested disproportionately more often than white youth, relative to their incidence of crime
(c) African American youth and white youth are being arrested at proportionally the same rate, relative to their incidence of crime
(d) There is not enough information to tell if any of the above are true

(7) This question is the same as (6), except "incidence of crime" is replaced with "share of the population."

To help you reason about the next question, here are some quotes from the rest of the Huffington Post article [38] your statistics came from:

> "African American youth represent 48% of all youth incarcerated for a drug offense in the juvenile justice system."
>
> "According to the National Survey on Drug Use and Health, among youths aged 12 to 17, the rate of current illicit drug use was 11.1% among whites, and 9.3% among African Americans [1]. In a previous year, the same survey found that white youth aged 12 to 17 are more than a third more likely to have sold drugs than African American youth [54]. *The Monitoring the Future* [emphasis added] Survey of high school seniors shows that white students annually use cocaine at 4.6 times the rate of African Americans students, use crack cocaine at 1.5 times the rate of African Americans students, and use heroin at the same rate of African Americans students, and that white youth report annual use of marijuana at a rate 46% higher than African American youth [37]."

(8) Which is most consistent with your calculations and the information above?
(a) White youth are being arrested disproportionately more often than African American youth, relative to their incidence of crime
(b) African American youth are being arrested disproportionately more often than white youth, relative to their incidence of crime
(c) African American youth and white youth are being arrested at proportionally the same rate, relative to their incidence of crime

(9) This question is the same as (8), except "incidence of crime" is replaced with "share of the population."