2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Paper ID #52427

Simulating Interdisciplinary Ethics in the Engineering Business through an AI Startup Game

Prof. Trystan Sterling Goetze, Cornell University

Trystan Goetze is the Sue G. and Harry E. Bovay, Jr., '36 Assistant Teaching Professor and Director of the History and Ethics of Professional Engineering. They grew up in Kitchener, Ontario, Canada, and attended the University of Waterloo (Honours Knowledge Integration with Joint Honours Philosophy, 2008–12), York University (M.A. in Philosophy, 2012–13), and the University of Toronto (M.A. in Humanities, Social Sciences, and Social Justice Education, 2013–14), before taking their Ph.D. in Philosophy at the University of Sheffield (2014–18). Before coming to Cornell, Prof. Goetze worked as a Banting Postdoctoral Fellow in Philosophy at Dalhousie University (2019–21), an Ethics and Educational Development Consultant for Ethically Aligned AI, Inc. (2021), and a Postdoctoral Fellow of Embedded EthiCS at Harvard University (2021–23). In their research, Prof. Goetze has written about moral responsibility, epistemic injustice, the ethics of computing technologies, and interdisciplinary ethics pedagogy. In addition to directing the Bovay Program in the History and Ethics of Professional Engineering, Prof. Goetze is also an affiliate faculty member in the Sage School of Philosophy. In their spare time, Prof. Goetze designs, runs, and plays tabletop roleplaying games.

©American Society for Engineering Education, 2026

Simulating Interdisciplinary Ethics in the Engineering Business through an AI Startup Game

Trystan S. Goetze
Bovay Program in the History and Ethics of Professional Engineering
Cornell University

Abstract

While case studies are widely used to teach ethical reflection in engineering, this activity has two shortcomings that may hamper the transferrability of learning in engineering ethics to the professional context. The first difficulty is that the cases themselves are static, not changing in response to student choices, as well as being somewhat abstracted from their own experience. The second difficulty is that the learning cohort frequently lacks a range of different disciplinary perspectives, which are needed in resolving complex ethical issues in professional engineering. This paper introduces these two challenges, develops a pedagogical model inspired by philosophy of education and games, and presents a simulation game set at a fictional AI startup company based on this model.

Keywords

engineering ethics; case study; simulation game; game-based learning; AI ethics; philosophy of education; pedagogical theory

1. Introduction

Case studies are a common learning tool in engineering ethics, with studies occasionally suggesting noting that they may be *the* most popular method [1, p. 47], [2, p. 1132]. Collections of case studies are a common feature in engineering ethics textbooks (e.g., [3], [4], [5]). And libraries of cases, such as that maintained by the National Society of Professional Engineers (NSPE) Board of Ethical Review, are available for an engineer to consider in their ongoing professional development [6].

At their best, case studies provide a clear picture of the stakes of ethical decision making in a professional engineering context, for both the engineers at the center of the dilemma and their corporate and public stakeholders. A good case study should help students understand the kinds of ethical challenges they may encounter in their professional lives and the connections between technical engineering decisions and their ethical responsibilities to the public, as well as provide them with practice in deliberating on these issues and making decisions that they can ethically justify to themselves and their peers. The aim of examining cases is to prepare students to make responsible decisions when they enter the professional engineering workforce.

However, case studies as traditionally presented are not without flaws. In §2, I start by highlighting two challenges for case study–based ethics instruction. The first, which I call the Phaedric Problem, is that case studies are static and distant from students' experiences. This makes the experience of engaging with even realistically described and complex case studies difficult to translate into the actual context of making ethical decisions in engineering. The second challenge, which I call the 2+ Cultures Problem, concerns the learning environment in

which case studies are analyzed—specifically, the challenge is a lack of disciplinary diversity in the cohort of learners themselves. If we accept that many of the most urgent problems faced by professional engineers are complex and require perspectives from multiple different sets of expertise, then engineering students must gain an appreciation for perspectives that come from disciplines quite different from their own, and must practice communicating with people from different disciplines on shared problems.

The rest of the paper describes how I worked to address these challenges in a unique educational opportunity that engages with a case study through a simulation game. §3 fills out the background by introducing the pedagogical approach I employed, which combines elements of Weston's model of teaching as staging, Nguyen's theory of games as an art form that shapes our agency, and Dewey's account of play and work in education. §4 describes the simulation game, where participants from multiple different disciplines took on the role of members of a team at an AI startup. As they played through the stages of the game, participants were presented with a scenario that presented ethically contentious choices to be made. To make a decision, they had to engage in conversation with their team about the different issues at stake. The scenario then would change in response to their choice (as well as random chance), prompting reflection on the outcome ahead of the next stage. This game was used as the primary activity of the September 2025 Sidney & Ann Braudy and Louis & Edith Manker Workshop on Business Ethics in Engineering at Cornell University. The Appendix contains the materials used in the session, reformatted for these proceedings. The final section of the body concludes with some reflections on the limitations of the simulation, the workshop for which it was developed, and the pedagogical model that inspired its design.

I have several aims in proceeding like this; as a result the paper is quite lengthy. Here is a reading guide, structured by the intersection of my goals and the reader's: My first goal is to showcase the simulation game itself. Readers interested mainly in the final product may wish to skip to §4 and the Appendix. My second is to explain why one might design an engineering ethics case study exercise in this manner. Readers interested in the motivation that comes from the context of engineering ethics education will find §2 most relevant. Readers who want to go deeper into the pedagogical theory that informs the design of the simulation game will find §3 worthwhile. What this paper lacks is a study of the effectiveness of this approach—hence why it is presented as a Case Study / Practice Paper rather than a more traditional contribution to the ASEE Conference—which I hope to explore in future work.

2. Two Challenges for Engineering Ethics Case Studies

In this section, I outline the two challenges which motivated the creation of the workshop described in §4. The first concerns the nature of traditional case studies themselves: even the best cases are static and unresponsive to student decisions, contributing to a distance felt between the student as objective analyst of the case and the ethical lessons contained within a richly described and realistic scenario. The second concerns features of the cohort of learners who engage in an engineering ethics case study analysis: typically, they are mostly or entirely engineering students, which limits the perspectives that can be brought to bear on complex problems, and misses out on an opportunity for engineering students to practice interdisciplinary collaboration.

2.1. Static Cases: The Phaedric Problem

The first challenge for engineering ethics case studies concerns features of the case study and the accompanying learning exercise. Martin, Conlon, and Bowe asked a group of instructors about their preferred features of case studies: "Instructors consider that case studies should be realistic, experiential, relevant, engaging, provocative, facilitated properly, including various stakeholders, integrating ethical alongside technical or legal considerations, based on real or real-time data and documentation" [1, p. 56]. But one difficulty highlighted by an interviewee is simply finding case studies that meet these desirable criteria, with one instructor noting, "I think the challenge is actually finding good case studies […] as well as finding the documentation to support it, because very often some of the information is not made publicly available" [1, p. 58].

In what way are the many and varied case studies available deficient? Bairaktarova and Woodcock note that while engineering ethics cases often have realistic complexity embedded within them, their content is often distant from students' personal experience by virtue of being "events that happened to someone else" (real or fictional). They further criticize the application of ethical reasoning students are encouraged to make for being too abstract, "based on an established code of ethics and intuition of what is right or wrong rather than on personal experience" [2, p. 1132]. Even with a good case study, the experience of engaging with it might lack vivacity to the engineering student, since they are usually in a position of offering an objective *post hoc* assessment with the benefit of hindsight and a set of specific principles to apply, algorithm-like, to a static scenario.

Here we begin to see the issue. A typical engineering ethics case study exercise is like a dissection. An item—perhaps a frog, or a case study about, say, the *Challenger* space shuttle disaster—is presented to be analytically cut open and analyzed against a standard set of criteria —a diagram and list of organs, a code of ethics or primer on moral philosophical theory. The experience may well be educationally valuable, but it remains an engagement with something *dead*—whether it be an animal or an event in an engineer's career. Whatever is learned through the lesson may be transferrable to the live context—of interacting with live animals or making ethical decisions in the workplace—but the experience itself lacks the richness of an engagement with a *living* creature or scenario, which responds to one's choices dynamically and often unpredictably. (That said, the higher stakes of interacting with something *truly alive* call for caution: we are right to avoid cruel vivisections or putting students' lives or careers in jeopardy in the classroom! The ideal is something closer to real life while removing the worst risks.)

I call this challenge the *Phaedric Problem* because it is essentially the same issue that Socrates raises for learning from the written word in Plato's *Phaedrus* [7]. Socrates says that while written texts can help to transmit knowledge or artistic works, they also come with a risk. An uncritical reader might simply learn to repeat what they have read, without understanding the deeper meaning and broader context, and without the ability to assess the text for its truth or beauty. And, according to Socrates, even a critical reader will be hampered in their ability to fully engage their faculties in interacting with a text, because the words themselves cannot respond to questions, and do not change in response to the reader's interpretations and decisions. The written word is dead; conversation is alive. Similarly, case studies as traditionally presented are static artifacts that are unresponsive to a student's questions or decisions. A student could learn the rules suggested by codes of ethics and moral theories, apply them to features of a case, and still not be prepared for the complexity and variability of real-life ethical decisions.

2.2. Interdisciplinarity: The 2+ Cultures Problem

The second challenge I want to mention is also related to the desire for realism when engaging with case studies, but it attends not to features of the cases themselves, but rather, to features of the *learners* and how the makeup of the student cohort shapes the learning environment. The kinds of problems many engineers are called upon to solve today are complex enough to require the combined expertise of multiple disciplines within engineering and beyond [8], [9], [10]. To this end, multiple studies and pilot programs have been conducted to investigate how to integrate interdisciplinary collaboration into engineering education, e.g., [11], [12], [13], [14]. I call this the *2+ Cultures Problem*, as a nod to Snow's famous 1959 lecture on the growing divide between the scientific and engineering disciplines, on the one hand, and the arts and humanities disciplines, on the other [15]. The "+" is an acknowledgement that there are divisions between more than just the two cultures Snow discusses—not just between the scientific and humanistic, but also between these disciplines and those taught in business schools, for example.

Moreover, I would argue that demands often placed on engineers through engineering ethics education necessitate greater training in and exposure to disciplines quite beyond the technical training which dominates most engineering programs of study. For example, in his engineering ethics textbook, Taebi repeatedly notes that engineers need to be aware of the legal and social context of their work [5], but a full appreciation of these perspectives is simply not something one can gain from a single course or even a series of interventions across the engineering curriculum. In order to fully discharge their duties to understand and consult with a variety of public stakeholders—as required in various paradigms of responsible technology development, such as value-sensitive design [16]—it behooves engineers to work with those who have expertise in social, legal, and other disciplinary contexts.

The business context in which engineering firms operate also requires yet another perspective and set of skills and knowledge which are not central to an engineer's education. This disconnect between engineering and business perspectives is perhaps most clearly illuminated in a famous anecdote from the loss of the space shuttle *Challenger*. The senior management at Thiokol, the engineering company contracted by NASA for the shuttle's rockets, frustrated by their engineers' opposition to launching the shuttle under conditions they expected to cause disaster, ruled that the question before them was a "management decision," not an engineering one. When only one of the senior managers, Robert Lund, the Vice President of Engineering, hesitated to vote with his colleagues to authorize the launch, the Senior Vice President, Jerry Mason, directed him to "take off his engineering hat and put on his management hat," swaying his vote in favor of the fateful launch [17, p. 316]. In other words, Mason instructed Lund to look at the launch authorization decision from the perspective of someone concerned with *all* of the risks to the company, including financial losses and loss of goodwill with their client, should they call off the launch, in addition to the risks to safety raised by the engineers, which were well-founded but not yet certain at the time.

Robison's comments on the *Challenger* case, in his introduction to engineering ethics, are worth quoting at length, as they illustrate the 2+ Cultures Problem for an individual engineer:

> What the *Challenger* example illustrates is an all too common problem. We are trained into a discipline and so learn to think in a certain way and carry that with us as we approach other problems… As necessary as it is to learn to think like an engineer to be an engineer, it can be an enormous disadvantage, as Mr. Lund

> discovered, not realizing that other disciplines embody other modes of thinking that give different results when put to a problem. Unfortunately, learning to think like an engineer, or like a physician, or like a lawyer is difficult enough. To succeed, we must train ourselves out of whatever had been our standard way of looking at a problem and then make the mode of thought of our discipline second nature to us… Yet if the aim is to train ourselves into a mode of thought that becomes second nature, it may seem paradoxical for us to try to get out of our own skins and come to see that we have been trained into a mode of thought that other disciplines may neither share nor understand… If we aspire to be the best in our chosen profession, we must learn to understand both the strengths and limitations of its natural mode of thought. Thinking about a problem in only one way is a recipe for disaster. [18, pp. 164–166]

Engineers, according to Robison, need to develop an appreciation for other disciplinary approaches to problems. He seems to suggest that it would be desirable for engineers to acquire the ability to switch between these perspectives as appropriate in order to best understand complex problems and how to address them. But he also acknowledges that it is difficult work to develop just the expertise for only *one* such perspective. Further, as noted by Raji et al. in a critique of ethics education in computer science, it is unrealistic to expect technologists to be "unicorns" who can master every disciplinary perspective that is relevant to responsible innovation [19].

I would suggest that it would be more fruitful instead for engineering students to practice working on a team with people who come from different disciplinary backgrounds on the same problem. But this is difficult in a context like an engineering ethics or computing ethics course, which in my experience attracts limited interest from outside the discipline named in the course title. And in contexts where the curriculum *requires* engineering and computing students to take the course, limited space and a need to serve the course's primary audience may crowd out students from other backgrounds. I have explored some of these issues in the context of computer ethics education before [20], but at a curricular level. In the present context, I am interested in how to address the 2+ Cultures Problem at the level of a specific educational opportunity.

3. Philosophical Inspirations

In this section, I briefly introduce some of pedagogical ideas that informed the design of the case study and simulation game which I developed in response to the challenges introduced in the previous section. These inspirations come from several philosophers' thoughts on education, unified by attention to the role of play. The first is Weston's pedagogical model of teaching as staging. The second is Nguyen's theory of games as the art of agency, and how playing games enables us to learn different ways of exercising our agency. The third is Dewey's account of play and work in education as connected ways of attending to activities in the classroom for their own sakes. I have articulated some similar ideas before in outlining the background to another game-based engineering ethics activity [21].

3.1. Teaching as Staging

Anthony Weston is something of a celebrity in philosophy teaching circles. A longtime professor at Elon University, he became known for a popular ethics and critical thinking textbook [22], and

for practicing and propounding an approach to teaching that he calls "staging." Following the idiom of King's distinction between the teacher as *sage on the stage* (lecturer) and *guide on the side* (facilitator of mostly self-directed learning) [23], Weston also refers to his model of the teacher as *impresario with a scenario*. He defines the method as follows:

> *Teaching as Staging* is regularly setting up classroom scenarios such that the dynamics of the scenario frame and drive most of the action and learning that follow. [24, p. 3]

The analogy Weston makes by referring to a teacher as an impresario and teaching as staging is to the context of theater. Like the impresario who organizes a theatrical performance for their players, the teacher's role is to design a scenario for the students. But unlike a playwright, who produces a script for actors to follow, a Westonian impresario takes a role more akin to the leader of improvisational theater in the Italian tradition of *commedia dell'arte*. A premise and set of characters are provided, as well as specific events the performance should incorporate, but the actors themselves determine how to play out the scenario to its conclusion. A well-produced dramatic scenario provides sufficient structure for the actors to follow with minimal intervention, which requires the impresario to think carefully about the goals of the scenario and how it can be structured to have an internal logic that unfolds naturally when participants engage with the scenario on its terms.

Weston defines a *scenario* as:

> some kind of happening or state of affairs organized on a pattern known to all and at least partially revealed at the start… [which] progressively develops or unfolds following an inner logic that orients but does not completely determine the direction of the situation's movement or final resolution… enacted in a specific setting, such as a classroom. [24, pp. 46–7]

For example, Weston describes a lesson he developed with astronomer Tony Crider for their co-taught course on life in the universe. In order to make the philosophical and practical challenges of searching for extraterrestrial intelligence come alive, students are told upon the start of class that someone out in the world is trying to find them, and it's their job to make contact. What follows is a scramble as small groups of students try to figure out what sort of message to look for, where to find it, and how to reply. Meanwhile, the "alien"—actually a compatriot of the teachers—is attempting the same process, knowing that the class is looking for them but lacking a direct line. Upon success, the class is introduced to the "alien" on a video call for a debrief [24, pp. 65–7]. While relatively simple in design, this scenario dramatizes all of the same sorts of problems faced by a group like the Search for Extraterrestrial Intelligence (SETI), and unfolds through a structure produced by the constraints of communications technologies.

Weston's reasons for favoring teaching as staging are rooted in a detailed critique of both lecturing and student-centered learning, which we need not rehearse in depth here. The basic problem, as Weston presents it, is that both of these familiar methods leave one side of the student–teacher relation too *passive*. Much of the empirical research shows that lecturing is ineffective in large part because the *students* are too passive. Meanwhile a thoroughgoing student-centered approach risks making the *teacher* too passive and unable to effectively direct learners towards achievement of learning goals. Teaching as staging, he argues, overcomes these

difficulties by requiring students to actively engage with the scenario to produce learning through their experience—unlike the passive role students take when listening to a lecture—and by requiring the teacher to put a good deal of thought into how the scenario is structured so as to direct students towards educationally productive ends—unlike the passive role for the teacher suggested by some approaches to student-centered learning [24, pp. 25–43].

Teaching as Staging is the basic inspiration for the case study activity I devised. In the next subsection, I describe the theory of games which informs the specific decision to create a simulation game as the structure of the scenario.

3.2. Games as the Art of Agency

Gamification is an approach to active learning that has been excitedly discussed by proponents such as McGonigal [25] and Walz and Deterding [26]. The basic aim is to introduce game-like elements into the learning process to make learning tasks more attractive to undergo, and more rewarding to complete. Gamification can be an engaging way to motivate students, but all too often it is reduced to a thin veneer, as when points, leaderboards, and badges simply take the place of traditional grades, honor rolls, and accolades. No one who is paying attention is fooled by this fresh coat of paint. To leverage the advantages of gamification under Weston's model of teaching as staging, we need a more sophisticated understanding of games to motivate gamification. For this, I turn to a theory developed by Thi Nguyen.

Nguyen argues that games are an art form that uses human agency as its medium. The game designer defines specific rules for the actions players can take and how those actions are resolved, as well as the goals players are aiming to achieve and the context in which the action occurs. The work of art—that is, the game—emerges when the game design is enacted in play, as players voluntarily submit to the constraints on their actions and desires for the duration of the game [27].

Just as we can take lessons learned from our engagement with other forms of art into our everyday lives, playing games can teach us about how to make use of our own agency. They allow us to experiment with taking different kinds of actions in various contexts and under the guise of various objectives, all of which we might not otherwise engage with. In so doing, by playing games we fill out a range of possibilities for how we exercise our autonomy:

> One game might be focused on high-speed reflexes, another on calculative look ahead, or on diplomacy and bargaining, or on manipulating alliances and shared incentives. Wide exposure to games can enhance the autonomy of agents by making them aware of alternative modes of agency… We can help to broaden each other's knowledge of agencies, help to develop out capacities to switch between modes of agency, and our ability to find the right mode of agency. [27, p. 78]

The lesson for using gamification as a framework for scenarios in engineering ethics is to design games that help students to practice specific uses of their agency that resemble real-life tasks and goals. Games provide a framework for specifying goals and details of a scenario which are similar to the circumstances a professional engineer may encounter and the objectives they may have, but simplified. The game's rules then require responding to that situation and set of goals in ways that resemble processes of responsible decision-making in a real-life scenario of that sort, and can govern the ways the situation changes in response to decisions and unexpected

factors. In the process, students can practice exercises of their agency in ways that more closely resemble being in those scenarios in real life, in a low-stakes environment.

But is this all just fun and games? To supplement Nguyen's theory of games and Weston's pedagogical model, let's add one more element that explains the educational value of play and how it bridges to the application of skills at work.

3.3. Play and Work

John Dewey wrote widely on the philosophy of education, since to him learning was at the heart of both individual experience of the world and the collective project of democracy. In an intellectual autobiographical essay written in 1930, he says that an interest in the theory and practice of education "brought together what might otherwise have been separate interests—that in psychology and that in social institutions and social life," such that his 1916 book *Democracy and Education* "was for many years that in which my philosophy, such as it is, was most fully expounded" [28, p. 156]. Here I want to focus on one small part of Dewey's vast corpus: the distinction between play and work in learning, how each has educational value, and how either can degenerate, if not properly structured, into pointless fooling or drudgery, respectively. This framework helps explain the educational value of a game-based Westonian scenario.

Anticipating the active learning movement by many decades, Dewey writes that students learn best when they are actively engaged in something they are independently interested in pursuing. What he calls *play* and *work* are both ways of achieving this: "Both involve ends consciously entertained and the selection and adaptations of materials and processes designed to effect the desired ends" [29, p. 202]. That is to say, both are activities where the student has to consider their goals and the means to achieve them, instead of simply being told a process to follow or being let loose without guidance. The difference between play and work has to do with the ultimate goal of the activity.

In play, the goal is the continuation of the activity itself. There are specific objectives to be achieved, but the point of trying to achieve those objectives is to play the game: "Persons who play are not just doing something… they are *trying* to do or effect something… The anticipated result, however, is rather a subsequent action than the production of a specific change in things" [29, p. 203]. Nguyen similarly observes that when we play a game, we experience a "motivational inversion" of our agency: "In ordinary practical life, we usually take the means for the sake of the ends. But in games, we can take up an end for the sake of the means" [27, p. 1]. In other words, in everyday life we choose activities in order to achieve our goals. When playing a game, by contrast, we are given a goal so that we can engage in an activity—namely, playing the game. The presence of specific goals and the conscious selection of means of achieving them is what distinguishes play from mere fooling. As Dennis explains in an essay recapitulating this aspect of Dewey's work, fooling is an activity that "is not self-directed nor is it imposed, but it is a matter of whim or accident. The sequence of acts in play is important, but there is no reference to sequence in fooling—the order is aimless and arbitrary" [30, p. 232]. Play has a structure that enables educational results to be embedded on purpose; fooling is unstructured and learning occurs by accident, if at all.

Work, in Dewey's sense, extends from play with a greater focus on achieving a specific end result. This requires more sustained effort and long-term planning around how to meet one's goals, whereas in play the focus is more near-term and aimed at continuing the activity rather

than a definite ultimate goal. But unlike the ordinary non-gameplay structure of motivation Nguyen describes, in Deweyan work the value of the activity itself remains primary: “Like play, [work] signifies purposeful activity and differs *not* in that activity is subordinated to an external result, but in the fact that a longer course of activity is occasioned by the idea of a result” [29, p. 204]. Engaging in work is still doing the activity because one is interested in the activity itself, but the activity becomes more serious, requiring greater structure and direction. What work is not is mere drudgery, where one loses interest in the activity because the goal and the means of achieving it are without value to the person engaging in the activity. As Dennis puts it: “labor is an activity carried on without any significance being attached to the actual doing, exclusive concern being with the result, which is often urged by some external pressure” [30, p. 232].

The lesson for us in creating a game-based scenario as a framework for engineering ethics education is to make use of the structure Dewey provides where play naturally develops into work. Providing the structure of a game at the start sets up a student’s motivation to engage directed behavior for the sake of play, without the risk of devolving into fooling. Goals are provided but the point of pursuing the goals is to play the game. As the game goes on and its internal stakes become more urgent, the student continues to engage in the activity for the sake of continuing the game, but the goals in view shift to include a higher level. The point is not just to successfully complete one stage of the game to move to the next, but to play the game as a whole in a particular kind of way—the activity changes from play to work. The design goal for this kind of exercise is to encourage students to progressively reflect on the trajectory of their game session as a whole, with an eye towards successfully simulating the actions of a responsible engineer. Because the student’s primary interest is continuing the activity for its own sake, however, it never degenerates into drudgery.

4. Simulating Ethical Issues at an AI Startup

In this section, I describe the workshop activity inspired by these different theories and commitments. I start by outlining how the theoretical background described in the previous section comes together to address the Phaedric Problem, and how I approached the 2+ Cultures Problem in organizing the event. I then describe the structure and flow of the game. The Appendix contains the actual game materials and facilitation guidance that was used.

4.1. A Model to Address the Phaedric Problem

First, let’s assemble the lessons of the last two sections into an explicit model of case study design. Recall that §2 argued that there are two challenges for case studies in engineering ethics: static case studies that are abstracted from students’ experience and unresponsive to decisions made in response to the case (the Phaedric Problem), and a lack of interdisciplinary expertise in the classroom (the 2+ Cultures Problem).

The pedagogical lessons from §3 primarily address the Phaedric Problem. From Weston, we get the notion of a scenario, a framework for the case study activity that requires participants to take an active role in the case itself, which is dynamically responsive to participants’ choices. Following Nguyen’s insights about games, the needed structure and dynamism are supplied by the rules of a game, which specify the goals participants have to achieve, the means they can use to achieve them, and the ways in which the scenario changes in response to their choices. This setup operationalizes Dewey’s account of how play and work are connected: with careful prompting to reflect on the broader stakes of one’s decisions and their outcomes within the

framework of the game, the initial playfulness of the game transforms into work by drawing attention to the wider stakes of the activity, while keeping participants' interest in continuing the activity for its own sake. We should also expect, following Nguyen, that engaging with a case study in this way will be more readily transferrable to the professional context, since one effect of playing games is to develop one's capacity to exercise one's agency in ways similar to one's experience playing a game.

4.2. Organizing an Event to Overcome the 2+ Cultures Problem

The final piece of the puzzle is interdisciplinarity. This is in part a logistical problem—one of getting the right mix of people into the learning environment—and could be tackled by finding ways of redesigning course admissions structures, opening up curricular constraints in engineering colleges, and collaborating with colleges of arts, business, law, and so forth to promote more enrollment in engineering courses by non-engineering students. But that approach would be Sisyphean, potentially requiring a radical redesign of engineering education institutions, or even universities as a whole. In this case, the solution is to work around the problem.

The way I addressed the 2+ Cultures Problem was to organize an extracurricular workshop advertised not just to engineering students but also explicitly to law and business students, as well as any students from other relevant disciplines who were interested in the workshop's theme. In total, 20 student participants attended the workshop, 10 of which came from engineering or computing. Facilitators from various different disciplines were also recruited, including faculty, an alumnus, and a donor. In total, 8 facilitators participated, 4 of which came from engineering or computing. During the workshop, students were assigned to groups that contained representatives from several different disciplines in order to foster cross-disciplinary conversation. The environment created as a result of this planning ensured that each group had a wide variety of perspectives to draw from, including several areas of engineering, computing and information science, law, business, philosophy, and science & technology studies.

The effort to make the workshop interdisciplinary was not entirely on my own initiative: some years ago, an alumnus of and donor to our engineering school saw the results of the 2+ Cultures Problem in his professional life. To address the consequent need for students to practice interdisciplinary interactions in tackling ethical issues in the engineering business, he created an endowed fund specifically to support such a workshop. While the workshop has existed for nearly two decades in some form, as we planned the most recent iteration of the workshop, the donor and I agreed on the outlines of the Phaedric Problem and the 2+ Cultures Problem as challenges we should address. The result is the simulation game described below and presented in the Appendix.

4.2. Outline of the Workshop

The structure of the workshop had three main phases: (1) a keynote lecture, (2) the simulation game, and (3) a discussion debrief.

4.2.1. Keynote

I delivered the keynote lecture to set the scene for the activity and discussion to follow. Building on the theme of "Artificial Intelligence and the Future of Life and Work," I introduced several

perspectives on the development of AI and how it is expected to impact our lives. I started by highlighting that what exactly people mean by "AI" is highly variable, which gives rise to wide-ranging disagreements about how the technology can be expected to progress, and how it will affect humanity along the way [31]. Drawing on Moor's account of computer ethics [32] and Vallor's theory of technomoral values [33], I then introduced the concept of an emerging technology as a new invention that changes what it is possible for us to do, making some things easier, some things harder, and some radically new things possible. To use emerging technologies wisely, I argued, we need to reflect on what they can do and how we can use them in accordance with our values. This set up further discussion of three sub-themes, which each raise their own questions for consideration:

1. ***Labor***: When should AI replace human labor, and when should it merely augment human labor?

2. ***Sustainability***: Is AI a sustainable technology—environmentally, socially, and economically—and if not, can we make it so?

3. ***Responsibility***: Who is responsible when AI makes—or just contributes to—a decision?

Emphasizing that none of these questions has clear and obvious answers, we then transitioned to the simulation game.

4.2.2. Simulation Game

The simulation game was played in groups of about 4 or 5 people, including a facilitator, whose role was to keep the discussion on track and on time, in addition to contributing their own perspectives. I acted as a master of ceremonies, circulating to answer questions, help locate materials, check in on progress, and move the whole group along through the stages of the exercise.

The game first introduces participants to the situation they will be playing through. They are all members of a team at a fictional AI startup, whose products up until now have been tailored AI chatbots for various purposes. The company has had several successful venture funding rounds and there are rumors that the company is approaching an initial public offering, which could benefit everyone who currently works there. The executives have promised their funders that they will diversify their product line, and have hired a polarizing figure as a chief AI officer, who will be the team's supervisor. This setup and the personality of the chief AI officer are based in part on several memoirs of working in Silicon Valley [34], [35].

Next, the game's procedure is explained. At each stage, there is a description of the current state of the situation: what the team's current challenge is, what pressures are coming from their boss, and what their options are. The group discusses the situation and their different possible choices, and brainstorms reasons for and against each choice. After 10 minutes, they come to an agreement on which option is best, and make a note on what they think the single best reason for that choice is. One person then rolls two six-sided dice and adds the numbers together. The result determines the nature of the outcome of their decision:

- On a roll of 6 or less, it is a bad outcome: some unforeseen event, bad luck, or mishap occurs, which may or may not be the team's fault.

- On a roll between 7 and 9, it is a good outcome: there is a positive impact from the team's decisions.

- On a roll of 10 or higher, it is an exceptionally good outcome: there is some additional benefit due to things going even better than expected. This result gives the team a small bonus to their next roll.

This dice-rolling procedure is adapted from the tabletop roleplaying game *Apocalypse World* and the diverse family of games—known as games "Powered by the Apocalypse"—that have been based upon it [36]. The nature of the result tells the group which card to retrieve from a pack; the card describes what happens next in the scenario, depending on whether they received a good, bad, or exceptional outcome. The narrative of the case thus changes for each group in response to both their choices and unexpected factors. This allows the situation to unfold through time in a particularized way and makes the participants' choices matter, while also simulating the uncertainty of actual decision-making by incorporating an element of chance from the dice roll. Finally, each outcome card also comes with a reflection question for participants to individually consider, raising further questions about their choice and its implications.

This procedure is repeated three times as the scenario evolves. In the first stage, the team needs to decide what kind of product to develop—the options are given as either a study aid chatbot for students or a regulation navigation chatbot for small businesses—and have to consider the risks and rewards of each option. In the second stage, they are asked by their boss to change the foundation model their product uses, and have to consider factors like the environmental sustainability of their options and the tradeoff between risks of hallucination when using a general-purpose model vs. the additional labor required to create a more tailored model that has a lower risk of hallucination. In the final stage, their product faces public scrutiny from a combination of two different groups of protestors. Their choices and their outcomes in previous stages determine which combination of protest groups they face and the nature of the criticism. For example, if they chose to make a study aid (and things went badly) with a large general purpose model, they encounter protest from a group of concerned parents and an environmental advocacy organization. Their choice at this stage centers on how (or whether) they choose to engage with this public criticism. This last stage takes inspiration from a simulation developed by the Scientific Citizenship Initiative for discussion of gene editing technologies, which includes a stage where a researcher is confronted by public protest [37].

4.2.3. Debrief

At the conclusion of the game, the whole group debriefed through discussion of several reflection questions. Some reflection questions were general in scope, asking about how the experience went or which choices were made. Other questions called back to themes introduced in the keynote. Discussion continued over catered dinner.

The purpose of the debrief was twofold. On the one hand, it was of interest to me as the workshop designer to hear from participants about their experience. On the other hand, these questions served to refocus on the learning objectives of the workshop, to share insights from the different pathways each group chose to follow, and to discuss any lingering questions that emerged from the experience.

5. Conclusion and Reflections

In this paper, I have presented some challenges for engineering ethics case studies as a pedagogical tool, outlined a model of game-based scenario-style teaching, and presented a case study workshop based on this model in order to address the aforementioned challenges. My initial concerns were that static case studies lacked relevance to students' experience, and insufficient interdisciplinarity in the classroom limited the relevant perspectives in discussion of cases. To address the first problem, I developed a model of active learning derived from Teaching as Staging, games as the art of agency, and Deweyan work and play. I thus arrived at a workshop design that invited participants from across different disciplines to engage in a game that requires participation in ethical decision-making in the context of an AI startup, structured through a fictional scenario that evolves in response to the choices made by participants, and which prompts reflection on the broader stakes of such decisions for a professional in the field. The result is a promising approach for more complex ethical case study exercises. However, the present work does have some limitations worth mentioning.

Potentially, the design of this game could be made more open-ended. As it stands, the design of the simulation has multiple paths but is still constrained in the choices available and the endpoints it reaches. It is perhaps closer to a script than a true scenario in Weston's sense. There is a great deal of variation still embedded in the design, but not the almost limitless possibilities of Weston's inspiration in improvisational theater. A more open simulation game would also more closely resemble a tabletop roleplaying game, like the one from which I took this game's dice mechanic. A possible tradeoff would be the need for skilled facilitation to keep the development of the scenario on track—not insurmountable in an arrangement that includes faculty facilitators in each small group, but this could be an impediment to scaling and adapting this lesson design to other contexts. On the other hand, some instructors may prefer the higher degree of structure and control that exists in the game as designed, as it ensures predictability in the events and lessons that can be drawn from them. While a Westonian scenario promises something similar from the internal logic of its design while remaining more open-ended in its development in play, this may also present risks of going to places that are both unexpected *and* unproductive.

While the workshop deliberately recruited interdisciplinary participants and created breakout groups with mixed expertise, one area that could be further developed in this regard is more explicit attention to interdisciplinary collaboration skills. As developed, the workshop somewhat takes for granted that students in different disciplines will have the skills needed to collaborate effectively across disciplines, but this perhaps should not be assumed. There are many different models of what interdisciplinarity is, how interdisciplinary collaboration should proceed, and how the underlying skills can be taught [38], which I have not yet incorporated into the workshop design. Taking the 2+ Cultures Problem seriously may, therefore, require more attention to the literature on interdisciplinarity.

Another area for further work is evaluation of the effectiveness of scenario-based case studies, whether or not they are structured through a game, as opposed to traditional case study analysis. Do such activities provoke more lasting insights about engineering ethics? Do students who engage in engineering ethics scenarios have greater self-efficacy when faced with ethical issues in a professional context? The answers to these research questions require further study. I find the theoretical model I have outlined and begun to put into practice compelling, but like everything else, it must be put to the test.

References

[1] D. A. Martin, E. Conlon, and B. Bowe, "Using case studies in engineering ethics education: the case for immersive scenarios through stakeholder engagement and real life data," *Australas. J. Eng. Educ.*, vol. 26, no. 1, pp. 47–63, Jan. 2021, doi: 10.1080/22054952.2021.1914297.

[2] D. Bairaktarova and A. Woodcock, "Engineering Student's Ethical Awareness and Behavior: A New Motivational Model," *Sci. Eng. Ethics*, vol. 23, no. 4, pp. 1129–1157, Aug. 2017, doi: 10.1007/s11948-016-9814-x.

[3] C. E. Harris, M. S. Pritchard, M. J. Rabins, R. W. James, and E. E. Englehardt, *Engineering Ethics: Concepts and Cases*, 6th ed. Boston: Cengage, 2019. Accessed: Jan. 13, 2026. [Online]. Available: https://faculty.cengage.com/works/9781337554503

[4] H. C. Luegenbiehl and R. F. Clancy, *Global Engineering Ethics*. Oxford, UK, and Cambridge, MA: Butterworth-Heinemann, 2017.

[5] B. Taebi, *Ethics and Engineering: An Introduction*. in Cambridge Applied Ethics. Cambridge: Cambridge University Press, 2021.

[6] National Society of Professional Engineers, "Board of Ethical Review Cases." Accessed: Jan. 13, 2026. [Online]. Available: https://www.nspe.org/career-growth/ethics/board-ethical-review-cases

[7] Plato, *Phaedrus*. London: Penguin Books, 2005.

[8] National Research Council, "Education for life and work: Developing transferable knowledge and skills in the 21st century," National Academies Press, Washington, DC, 2012.

[9] National Academy of Sciences, "Facilitating interdisciplinary research," National Academies Press, Washington, DC, 2004.

[10] National Academy of Engineering, "The engineer of 2020: Visions of engineering in the new century," National Academies Press, Washington, DC, 2004.

[11] L. R. Lattuca, L. C. Trautvetter, S. L. Codd, D. B. Knight, and C. M. Cortes, "Working as a Team: Enhancing Interdisciplinarity for the Engineer of 2020," presented at the 2011 ASEE Annual Conference & Exposition, Jun. 2011, p. 22.1711.1-22.1711.26. Accessed: Jan. 14, 2026. [Online]. Available: https://peer.asee.org/working-as-a-team-enhancing-interdisciplinarity-for-the-engineer-of-2020

[12] A. van den Beemt, M. MacLeod, and J. van der Veen, "Interdisciplinarity in tomorrow's engineering education: 48th Annual Conference on Engaging Engineering Education, SEFI 2020," *Engag. Eng. Educ.*, pp. 61–74, 2020.

[13] A. Van den Beemt *et al.*, "Interdisciplinary engineering education: A review of vision, teaching, and support," *J. Eng. Educ.*, vol. 109, no. 3, pp. 508–555, 2020, doi: 10.1002/jee.20347.

[14] G. P. Harrison, D. Ewen Macpherson, and D. A. Williams, “Promoting interdisciplinarity in engineering teaching,” *Eur. J. Eng. Educ.*, vol. 32, no. 3, pp. 285–293, Jun. 2007, doi: 10.1080/03043790701276775.

[15] C. P. Snow, *The Two Cultures*, Reissue edition. Cambridge: Cambridge University Press, 2012.

[16] B. Friedman and D. G. Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press, 2019. [Online]. Available: doi.org/10.7551/mitpress/7585.001.0001

[17] D. Vaughan, *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press, 1996. Accessed: Jan. 14, 2026. [Online]. Available: http://archive.org/details/challengerlaunch0000vaug

[18] W. L. Robison, *Ethics Within Engineering: An Introduction*, 2nd ed. New York: Bloomsbury, 2024.

[19] I. D. Raji, M. K. Scheuerman, and R. Amironesei, “You Can’t Sit With Us: Exclusionary Pedagogy in AI Ethics Education,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, in FAccT ’21. New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 515–525. doi: 10.1145/3442188.3445914.

[20] T. S. Goetze, “Integrating Ethics into Computer Science Education: Multi-, Inter-, and Transdisciplinary Approaches,” in *SIGCSE 2023*, Toronto, ON: ACM, 2023, pp. 645–651. doi: 10.1145/3545945.3569792.

[21] T. S. Goetze, “Roleplaying Game-Based Engineering Ethics Education: Lessons From the Art of Agency,” presented at the 2024 American Society for Engineering Education St. Lawrence Section Annual Conference, Toronto, ON, 2024. doi: 10.18260/1-2-170.1114.1120.1135.1150.1153-48578.

[22] A. Weston and B. Fischer, *A 21st Century Ethical Toolbox*, 5th ed. New York: Oxford University Press, 2023.

[23] A. King, “From Sage on the Stage to Guide on the Side,” *Coll. Teach.*, vol. 41, no. 1, pp. 30 –35, 1993.

[24] A. Weston, *Teaching as the Art of Staging: A Scenario-Based College Pedagogy in Action*. New York & London: Routledge, 2019.

[25] J. McGonigal, *Reality is Broken: Why Games Make Us Better and How They Can Change the World*. New York: Penguin Books, 2012. Accessed: Jan. 14, 2026. [Online]. Available: https://www.penguin.co.uk/books/406640/reality-is-broken-by-jane-mcgonigal/9780099540281

[26] S. P. Walz and S. Deterding, Eds., *The Gameful World: Approaches, Issues, Application*. Cambridge, MA: MIT Press, 2014.

[27] C. T. Nguyen, *Games: Agency as Art*. Oxford and New York: Oxford University Press, 2020.

[28] J. Dewey, "From Absolutism to Experimentalism," in *The Later Works of John Dewey, 1925-1953. Volume 5: 1929-1930, Essays, The Sources of a Science Education, Individualism, Old and New, and Construction and Criticism*, J. A. Boydston and K. E. Poulos, Eds., in The Collected Works of John Dewey, 1882-1953. (2nd Release). , Carbondale and Edwardsville, IL: Southern Illinois University Press, 2003.

[29] J. Dewey, *Democracy and Education*. New York: The Free Press, 1916.

[30] L. Dennis, "Play in Dewey's Theory of Education," *Young Child.*, vol. 25, no. 4, pp. 230–235, 1970.

[31] W. D. Heaven, "What is AI?," MIT Technology Review. Accessed: Jul. 18, 2024. [Online]. Available: https://www.technologyreview.com/2024/07/10/1094475/

[32] J. H. Moor, "What is Computer Ethics?," *Metaphilosophy*, vol. 16, no. 4, pp. 266–275, 1985.

[33] S. Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press, 2016. doi: 10.1093/acprof:oso/9780190498511.003.0012.

[34] G. Lewis-Kraus, "One Startup's Struggle to Survive Silicon Valley's Gold Rush," *Wired*. Accessed: Jan. 16, 2026. [Online]. Available: https://www.wired.com/2014/04/no-exit/

[35] D. Lyons, *Disrupted: My Misadventure in the Start-Up Bubble*. New York: Hachette Books, 2017.

[36] D. V. Baker and M. Baker, *Apocalypse World*, 2nd ed. 2016.

[37] Scientific Citizenship Initiative, "CRISPR ONLINE LEARNING MODULE." Accessed: Jan. 10, 2026. [Online]. Available: https://scientific-citizenship.org/crispr-simulation/

[38] R. Frodeman, Ed., *The Oxford Handbook of Interdisciplinarity*, 2nd ed. Oxford: Oxford University Press, 2017. [Online]. Available: https://doi.org/10.1093/oxfordhb/9780198733522.001.0001

Appendix: Simulation Game Materials

The following pages contain the materials used for the simulation game, including:

- A description of the fictional background for the scenario
- Instructions to participants
- Each stage of the scenario, with possible outcomes of each choice and reflection questions
- Debrief questions

Scenario Background

A.I. and the Future of Life and Work

You and your breakout group are going to imagine that you are a team working for a tech startup called JinnAI (a name punning on "jinni," a kind of spirit from Arabian myth that is the origin of stories of wish-granting genies).

The startup's focus is on creating tailored AI chatbots for various purposes. The founders pitched the company as an intermediary between general-purpose chatbots and models, on the one hand, and the consumer, on the other. The company's services include fine-tuning chatbots for specific needs and contexts. So far, the company's products have mostly been customer service chatbots for online businesses.

JinnAI has had several successful venture funding rounds. There are rumors that the company is approaching profitability and may soon launch an initial public offering (IPO), which would enable all of you to take advantage of your stock options, if the share price is high enough then. In the last funding round, the executives pledged to their investors that the company would begin diversifying its offerings, since customer service is becoming a saturated market for chatbots. They also stated that they intend to make the fullest possible use of AI within the company as well, in order to find efficiencies and maximize productivity.

With these goals in mind, they have hired August Blackwood as the company's first Chief Artificial Intelligence Officer (CAIO). When Blackwood came on board, he announced to the company that he's excited to join JinnAI as it makes the pivot to being, in his words, "AI-first, AI-optimized, and AI-led." His qualifications come from his computer science degree and previous work in venture capital firms that fund machine learning developers. Your team will be reporting to him for your next project.

Procedure

This simulation will progress through several stages as you make ethical decisions and the scenario evolves. To facilitate this, we will use some rules inspired by tabletop roleplaying games. Each stage of the simulation will proceed as follows:

1. Obtain the Scenario Sheet for the stage of the simulation you have reached. Read the description of the scenario for this stage and the decision your team needs to make. This should take about 2 minutes.

2. As a group, discuss the provided options for your decision. Brainstorm reasons for and against each choice, focusing on ethical reasons rather than economic reasons. This should take no more than 10 minutes.

3. As a group, come to an agreement on which option is best and make a note of the most convincing reasons supporting this choice. This should take about 5 minutes.

4. One person rolls two six-sided dice (2d6) and adds the numbers together. If you have a Token, add 2 to the result, then discard the Token. Finally, obtain and read the Outcome Card for this stage as follows:

- If you rolled 6 or less, take the Bad Outcome card for this stage. This represents some unforeseen event, bad luck, or mishap occurring after your decision—it may or may not be your fault.
- If you rolled between 7 and 9, take the Good Outcome card for this stage. This represents good results from your decision that have a positive impact.
- If you rolled 10 or higher, take the Good Outcome card and a Token. This represents some additional positive outcome from your decision due to things going exceptionally well.

5. Individually, reflect on your decision in light of the outcome. Use the prompt on the Outcome Card as a starting point. This should take about 2 minutes. Return to step 1 for the next stage of the simulation.

Scenario Stages

Stage 1 — Product Development

As part of JinnAI's move to diversify it sproducts, Blackwood (the CAIO) leads a series of brainstorming sessions with the various development teams. The meetings have a lot of excitement and creativity. Morale is high given the rumors that the startup is gearing up to go public.

You notice something odd, though. Blackwood has two distinct ways of communicating with his teams, almost like you're reporting to two different people. In-person, he is a helpful critic of project ideas, particularly around how to make ideas more marketable. He makes confident and clear decisions about specifications and timelines, and has a wry sense of humor. But over email and chat, he is much more vague and less critical. While generally more cheerful in messages than in-person, he's also less funny. His secretary tells you (off the record) that Blackwood is increasingly using an AI application to compose his emails and chats. She finds this habit frustrating, because it means she has to spend time chasing him for a real reply when the AI misunderstands, when she would rather spend the time answering routine company emails herself, as she used to do for the CEO at another company.

You don't have much time to ponder why the CAIO is eager to have a bot handle company emails for him, though. Your discussions have narrowed down the various options for AI products you could make to two. Both would leverage your existing technology by using a chatbot interface to access materials from a curated database and information contained in the model.

Your options are:

Option 1 — StudyBot . StudyBot would be a series of chatbot tutors and study aids, tailored for support in different subjects. Customers could be students who pay for individual subscriptions, or institutions like schools or colleges that buy a licence for all of their students. With schools often short on staff and resources, and students under ever-greater pressure to succeed, StudyBot could be an appealing supplement to existing educational support.

Option 2 — Regulation Navigator. Regulation Navigator (RegNav for short) would be a chatbot interface for small businesses to explore government regulations that apply to their situations. It

would offer help finding relevant laws and rules with explanatory guidance, while explicitly not being legal advice. RegNav would help fill a need for small business owners operating in an increasingly complex regulatory environment, saving them a great deal of time and money.

Outcomes for Stage 1

Stage 1 — Option 1 — Bad Outcome

You start with a focus on a bot that offers tutoring in STEM subjects at a pre-college level. It works okay, but has a bias towards older and outdated materials, particularly in human biology. This frustrates some users in politically progressive communities, whose curricula have more LGBTQ-inclusive discussion of sex and gender. It becomes difficult to balance meeting the progressives' desires with those of more conservative clients, who oppose the very changes the progressives desire. Debate erupts internally at JinnAI about how to respond.

Reflection: How might the company balance the conflicting values and desires of its different clients and its internal staff when iterating on this project?

Stage 1 — Option 1 — Good Outcome

You start with a focus on a bot that offers tutoring in STEM subjects at a pre-college level (ages 13–18). It tests well, and attracts interest from students and teachers. Some parents buy their kids subscriptions, and a few schools begin negotiating with the company to purchase institutional subscriptions.

10+ Bonus (take a token): A teen goes viral after sharing a video of a study session with the bot, showing the chatbot outperforming a tutor their parents had hired. Free marketing!

Reflection: What ethical issues are there with a product which is aimed primarily at providing an educational service to minors, and how should that be taken into account when developing something like StudyBot?

Stage 1 — Option 2 — Bad Outcome

You pilot RegNav in small city with relatively complex regulations, offering trial periods to small businesses located there. The bot performs well at first, and business owners are happy about the money and time they save. However, you start receiving a large influx of customer service calls when these businesses work on their local tax obligations—and they have questions you aren't qualified to answer. In a few complex cases, businesses complain that the bot misled them into making errors that have resulted in costly fines.

Reflection: How much responsibility do JinnAI or you, as the designers of RegNav, bear for the sometimes confusing and occasionally harmful consequences of using this product?

Stage 1 — Option 2 — Good Outcome

You pilot RegNav in small city with relatively complex regulations, offering trial periods to small businesses located there. The bot performs well and rarely hallucinates. Business owners happily report that you're saving them a lot of time and money.

10+ Bonus (take a token): A city official reaches out to ask if you could integrate RegNav into the city website, as part of the government's outreach programs.

Reflection: It sounds like the business owners are willing to put a lot of trust in your product. How wise is that? Would you recommend that they still consult with a legal professional on regulatory compliance? Why or why not?

Stage 2 — Foundation Model Refresh

Some time after the push to diversify the company's products, Blackwood again tells the development teams that it's time to shake things up. This time, he wants to reconsider the foundation models that the company uses. While this would be disruptive to the ongoing project that you're working on, it could be an opportunity to adopt a foundation model that better supports the tasks you need it to perform.

After much research and testing, you determine that there are two options that will work well for your purposes. However, each has some downsides as well.

Your options are:

Option 1 — BigNet . BigNet is a very large foundation model which is widely used and well-received. The purposes it is advertised for are very general: the model vendor suggests that just about any kind of text generation is a suitable task. Because of BigNet's size, it is impossible to run locally on your company's own computers. Adopting this model would thus make JinnAI dependent on the model vendor's cloud service. An advantage of using the cloud service is that the model vendor has more capacity to expand as your customer base grows, while your own company's machines are limited by funding and physical space. BigNet also has a reputation for being relatively easy to work with, generally requiring less trial and error to fine-tune. However, its popularity means that the model vendor's servers are sometimes the target of cyberattacks, which could bring security risks. BigNet is also prone to hallucination (e.g., generating falsehoods) if not prompted carefully, because of its high generality.

Option 2 — GreenGPT. GreenGPT is a more efficient foundation model that is a favorite of smaller developers. This model is also more suited to the specific tasks your product performs, reducing the likelihood of hallucination once properly fine-tuned. The model is small and efficient enough that it can even be run on your company's own servers, giving you greater control over customer data without the intermediary of a cloud service. However, growth could be constrained since new hardware would be very expensive. Your product could also be less reliable due to the need for scheduled maintenance and the possibility of power disruptions in your area—service interruptions which a cloud service smooths out by using multiple sites. The "Green" in the model's name is a reference to the model's efficiency, which means that it required considerably less energy to train and needs less energy to run than many of its competitors. However, GreenGPT requires a lot more work to fine-tune, and is much more difficult to update without having to redo much of the fine-tuning work, making products based on it less able to adapt to changing information.

Outcomes for Stage 2

Stage 2 — Option 1 — Bad Outcome

From the standpoint of technical fit, BigNet is a success, and leads to functional improvements for your product. It is a struggle to keep up with increasing cloud computing costs, but you just manage it. However, you run into a problem with de-biasing as you fine-tune the model. For some reason, the model has a tendency to give less helpful responses if the user introduces themselves to the chatbot with a feminine name, like Stacy or Lilly. The model vendor admits that this is a limitation that they're also having difficulty fixing. They believe it is due to a pattern in the training data available to them.

Reflection : If a client were to discover the bias introduced into your product by BigNet, how fair would it be for them to accuse your product of sexism? How much blame, if any, should you accept for this? Why?

Stage 2 — Option 1 — Good Outcome

BigNet works out well. The growing cost of the subscription to cloud computing is counterbalanced by the increased efficiency of your work, since you don't need to spend a lot of time fine-tuning with every model update.

10+ Bonus (take a token) : Your team wins a developer award at a conference hosted by OpenB, the company that owns and maintains BigNet.

Reflection : Research is inconsistent, but some estimates suggest that training a large model like BigNet emits carbon equivalent to a cross-country flight, and every single prompt can require up to half a liter of water for cooling. By using BigNet, are you responsible for contributing to the climate crisis? Why or why not?

Stage 2 — Option 2 — Bad Outcome

GreenGPT enables you to have efficient, in-house architecture for your product. However, as JinnAI looks to expand its client base, it becomes challenging to maintain your product. You end up competing with other teams for servers and compute power, as everyone tries to prove the merits of their work with limited hardware. The CAIO tries to spin this as healthy competition that will allow the best ideas to rise to the top. But practically speaking, it means that while some of your early clients have reliable access to your product at all times, newer clients complain about slowdowns and errors during peak hours.

Reflection: Is JinnAI's approach to sharing computing resources fair to its development teams or to its customers? Why or why not?

Stage 2 — Option 2 — Good Outcome

While it does take more work to fine-tune GreenGPT, you're able to anticipate this and compensate for it in your project management. Your success convinces JinnAI to acquire more servers and GPUs before they're needed, avoiding an internal scramble for scarce computing power.

10+ Bonus (take a token) : Your product wins a developer award at a sustainable computing conference.

Reflection : Your choice of model saves a lot of electricity, but it also requires JinnAI to acquire expensive hardware that is made with rare earth elements and specialty plastics. These components produce a great deal of toxic waste when extracted from the earth, and are difficult to recycle. So, how sustainable is your product, really?

Stage 3 — Opposition

Several months and iterations on your product later, leadership is pleased with your work and looking forward to a successful IPO. But things soon become complicated as two groups start raising public concern about your product.

The nature of the protestors depends on previous events. Retrieve two Protest Cards as follows:

- Stage 1
  - StudyBot
    - ► Good Outcome → Protest Card 1
    - ► Bad Outcome → Protest Card 2
  - RegNav
    - ► Good Outcome → Protest Card 5
    - ► Bad Outcome → Protest Card 6

- Stage 2
  - BigNet → Protest Card 3
  - GreenGPT → Protest Card 4

Read both of your Protest Cards before proceeding.

Before you can act, Blackwood puts out a statement that mostly ignores the concerns of the protestors. The statement focuses instead on how exciting it is that the "AI Revolution" has attracted opposition, and grandstands on that theme. The other executives are concerned about how this might affect JinnAI's upcoming valuation and ask you to consult with the company's marketing and outreach office to come up with a strategy.

Your options are:

Option 1 — Don't Engage . You decide that engaging with the protestors is too risky. Especially with the IPO approaching, negative press needs to be minimized. Your team will not speak with the protestors, and JinnAI will make no further public statements.

Option 2 — Sympathetic, Direct Response . You decide that you want to engage with the protestors to hear out their concerns and potentially incorporate their ideas into future iterations of your product. The outreach team will arrange for you to speak with their leaders.

Option 3 — Reasoned, Indirect Response . You think your product is worth defending against the protestors' concerns. With the outreach team's help, you will draft a formal reply to the protestors' concerns, explaining why they are wrong and refuting their arguments.

Protest Cards

Protest Card 1 — Educators

CHILL (Coalition for Human Intelligence in Lessons and Learning) is a group of teachers, advocating for "authentic education" which means, to them: no screens (except in computer class), no AI, small classrooms, community experience, and hands-on work. Their membership includes a diverse group of schools whose interests are not usually aligned, including alternative pedagogies (e.g. Waldorf, Montessori), religious schools (including Christian, Jewish, and Buddhist worldviews), schools serving low-income neighborhoods, and a few private liberal arts colleges. The founder is a leading scholar of the cognitive science of teaching and learning at a major teacher's college. CHILL's concern with your product is that, according to them, it undermines education by providing a shortcut to doing homework. They cite evidence that students are performing worse on tests when they can't use the AI tutor. These students tend to come from schools that are under-resourced and rely on tech companies for materials and services.

Protest Card 2 — Parents

A group of concerned parents begins protesting your product because they find the changes it's making to education troubling. They are worried that your product is part of a trend where students get less time learning from human teachers and forming bonds with each other. More concerning from a legal perspective is their claim that a number of students are using your product under false credentials. Because of strict data protection laws that apply to young children, JinnAI chose to limit access to users aged 13 and up. But these parents allege that many children are lying about their own age so they can sign up for an individual account. They also claim that some administrators at under-resourced schools are lying about the age of their students—claiming to be post-secondary schools when they actually teach grades 1 to 6—in order to get access, since it is cheaper than hiring more teaching staff. Sensitive information about minors has thus found its way into the data held by your company.

Protest Card 3 — Environmentalists

An environmental advocacy organization has launched a campaign opposing the use of BigNet, which they have turned into a poster child for, in their words, "the environmentally destructive excesses of AI." Your product is one of several they have singled out for its reliance on BigNet as a foundation model. One of their arguments pointing to the concrete impact of Big Net is that OpenB, the model vendor, has strained the electric grid to the point where they needed to purchase supplemental power. They have even reopened several gas-fired power plants that would otherwise have been phased out in favor of renewable energy, since natural gas power is easier to start up and more consistent in output. Another argument is that the amount of water they use to cool their datacenters is insultingly wasteful, given that the suburbs near those same datacenters are under water rationing due to a drought scientists attribute to climate change. Altogether, they conclude that it is unconscionable for anyone to work with BigNet.

Protest Card 4 — Labor

Your product has been identified by a labor advocacy group as, in their words, "replacing human workers with AI trained on those same workers' labor." They argue that automating their work

devalues their labor and creates a workplace environment where they are unlikely to receive fair compensation. They also believe that it adds insult to injury that the training data in your foundation model comprises, in significant part, work created by the very same laborers they represent. Most of the workers whose interests this group represents are not unionized; they claim that their advocacy is based on the notion that labor rights in the current economy require going beyond traditional collective bargaining methods. They add that even if these workers were unionized, the increasing use of AI to automate work would compromise the effectiveness of strike actions and the leverage that laborers have in negotiations, since their value is in question if an AI tool can do even some of their work.

Protest Card 5 — Legal Professionals

The local bar association issues a statement criticizing RegNav. They argue that there can be no substitute for a trained legal professional when seeking legal advice on any given topic. They argue that it's one thing to know what the laws are, it's another to understand how they apply to your situation. RegNav only locates potentially relevant parts of the law; it can't properly interpret and apply the law. While you make such disclaimers in the product's documentation, the bar association claims that business owners are ignoring your guidance and ending up in legal trouble. The bar association thinks that your product is, therefore, reckless, and risking a future lawsuit. They also argue that products like RegNav undermine trust in the legal profession. The profession exists because the general public needs highly trained people to assist with understanding and applying the law. Because your product is much cheaper and gives answers that seem satisfactory, it leads clients to question the value of a trained legal professional.

Protest Card 6 — Regulators

Regulators at a city where you recently deployed RegNav issue a public awareness statement warning that your product is an unreliable source of information. They strongly encourage business owners to consult with municipal clerks or legal professionals instead. They cite several instances where business owners got into complicated legal trouble because they uncritically followed your bot's advice. Around the same time, a reporter for a local paper publishes a story revealing that RegNav has helped a group of local businesses discover and exploit loopholes in the municipal regulations. Following this advice allows a business to avoid certain taxes, environmental assessments, and other obligations—all legally. The city is actively combing through their legislation to try to close these gaps, but in the meantime they want businesses to stop, calling it acting in bad faith. The reporter also notes that the city has already closed several loopholes, but at time of writing RegNav still recommends exploiting them.

Stage 3 Outcomes

Stage 3 — Option 1 — Bad Outcome

The protestors are incensed at JinnAI's silence. With nothing else to work from, they dunk on Blackwood's response for more than a week. Eventually, he deletes the post, but the story goes viral again after a prominent AI-skeptical content creator reposts a takedown of his remarks. For weeks, JinnAI is blasted on social media for appearing insensitive to the public. The company's valuation is negatively affected, and there is real concern that you and other employees might not get to exercise your stock options at the IPO.

Reflection : Despite choosing not to engage with the protestors, do you think they have a point about any of their contentions? If you were to respond to them, what might you say?

Stage 3 — Option 1 — Good Outcome

With no response from you to keep attention focused on your company, the protest moves on to other organizations producing similar products. The valuation goes well, and you expect to make a decent, but not life-changing, amount of money at the IPO.

10+ Bonus : You're still able to address some of the protestors' concerns in an update, which the CAIO spins as listening to stakeholders. The IPO goes very well: you make enough to take a year off—Blackwood says that he expects to retire early.

Reflection : How did the conflict between JinnAI's economic interests and your ethical values play out in this decision? Did you feel the conflict, or did you find alignment?

Stage 3 — Option 2 — Bad Outcome

The discussion with the protestors is consistently difficult, and you have a hard time building trust and common ground. The CAIO's statement makes things worse, particularly after he refuses to apologize at a group meeting. The ongoing controversy affects the company's valuation; there is real concern that you and other employees might not get to exercise your stock options at the IPO. Reflection If you were to try again, where might you be able to find common ground with the protestors? What concerns or values do you share that you could use as a starting point?

Stage 3 — Option 2 — Good Outcome

After several conversations with the protestors, you bring the temperature down and learn to trust each other. You start developing changes in response to their concerns. The product is better for it. The valuation goes well, and you expect to make a decent, but not life-changing, amount of money at the IPO.

10+ Bonus : The community-driven development work gets a favorable writeup in a news magazine, which attracts a surge of customers. The IPO goes very well: you make enough to take a year off—Blackwood says that he expects to retire early.

Reflection : There was a real risk that engaging could make things worse. How do you feel about the role of luck in the ethical impacts of technical and business decisions?

Stage 3 — Option 3 — Bad Outcome

Your response irritates the protestors and they double down in their actions. A small group of picketers assembles every day for a week outside the office building where your company is based, and the landlord is annoyed. Blackwood says you should have left it to him. The ongoing controversy affects the company's valuation; there is real concern that you and other employees might not get to exercise your stock options at the IPO.

Reflection: Sometimes in a contentious dispute, it's necessary to build common ground by acknowledging the other party's feelings, even if those feelings come from different opinions on matters of fact. With this in mind, what might be a more effective starting point to responding to the protest?

Stage 3 — Option 3 — Good Outcome

The protestors respect that you took the time to respond, and switch to seeking regulatory change instead of pressuring you directly. The valuation goes well, and you expect to make a decent, but not life-changing, amount of money at the IPO.

10+ Bonus : The CEO and CAIO use your arguments to persuade state and national regulators to create policy favorable to JinnAI. The IPO goes very well: you make enough to take a year off—Blackwood says that he expects to retire early.

Reflection: Look again at the protestors' claims. When you take JinnAI's success and the protestors' concerns into account, what is your honest judgement of how fairly the benefits and burdens of AI development are distributed across society?

Debrief Discussion Questions

- General
  - How did the simulation go?
  - Were there decisions that you found particularly challenging?
  - Were there decisions that you found easier than others?
  - Are there any outcomes that you feel uneasy about? Or really good about?

- Values
  - Which personal values were most important to particular decisions?
  - Were your values ever in tension with each other or with the values of the company (or your boss)?
  - How did you prioritize or reconcile these values when they were in conflict?

- Public Engagement
  - How did you handle the protest?
  - Is it the responsibility of companies developing emerging technologies to engage with stakeholders?
  - How should they respond to public disagreement?

- Themes
  - After the simulation, what are your thoughts on the three themes I introduced?
  - When is it appropriate for AI to replace, rather than just enhance, human labor?
  - Is AI a sustainable technology? If not, how can we make it so?
  - When AI makes (or contributes to) a decision, who is responsible?