Application for Mozilla Responsible Computing Challenge

Proposal by Arvind Narayanan and Steven Kelts, along with partners at The Princeton Center for Information Technology Policy and the non-profit AI4All

Title:
Agile Ethics:
Teaching the Ethics of Computing with Agile Methods

_____

Project Purpose (140 characters or fewer)

To develop role-play modules on responsible computing, available to enhance college courses, and supplement access/inclusion outreach. (133 characters)

Approach (methodology) (2500 characters or fewer)

Ethics of Computing (Fall '23) will be Princeton's first Computer Science course dedicated to social and moral dilemmas in tech. We are Computer Science Professor Arvind Narayanan (lead instructor) and Public Policy Lecturer Steven Kelts (an Ethicist affiliated with our Center for Human Values).

     This grant will support the improvement of a suite of "Agile Ethics" simulations, for use in this course. In these simulations, students inhabit roles on an Agile team (engineers, project managers, etc.) to experience ethical dilemmas as they emerge in real tech work. The students make technical and design choices, then discover through successive "sprints" that their choices may violate ethical norms. The role-plays also introduce basic principles from academic Ethics, like harms, duties, virtues, justice, structural inequity, care, etc.

     Each role-play will dovetail with a topic covered in "Ethics of Computing." As the students learn about technical design, they will role-play design processes for the same tech. Some Agile Ethics role-plays were piloted this year with an internal grant; the effects of the pilot are being studied by a Neuroscience team on a grant from a large tech company. The program was developed in consultation with responsible innovation professionals in industry – we invited tech professionals from NYC to share their knowledge, representing the diversity of those working in the space. As we learn more about our program's success, it

will be used with engineering teams in Silicon Valley and made available widely to universities, high schools, and nonprofits.

The role-plays are designed to give students exposure to real work processes and design principles, as well as the emergent nature of technmoral dilemmas. Ethical problems can be unforeseen, latent in early project choices. But standard case studies that highlight already-known harms – often from well-publicized examples – may not prepare students for real-world conditions of unanticipated wrongs. Ethics classes may fare no better, since they focus on theories of moral reasoning which only aim to improve judgment ***after*** a dilemma is recognized. So the Agile Ethics role-plays begin with recognition. They target students' "Moral Awareness," as defined in Rest's 4-stage psychological model of moral decision-making: our pilot study aims to confirm this effect. Instruction in Ethics then encourages the students through the latter stages: judgment, intent-formation and action. (2555 characters)

Deliverables (1250 characters or fewer)

Mozilla funding will be used to gamify the Agile Ethics role-plays in a mobile-ready form, to facilitate their widespread use. We will:
1) Design platforms for hosting the role-plays (including remote play), for facilitating participant input, and for scoring participation – compatible both with major learning management systems and independent third-party software.
2) Produce 20+ role-plays on different technologies, aiming for broad enough coverage to encourage adoption both in other Princeton courses with interested faculty (COS 226, COS 324, COS 432, and courses in Engineering fields) and an array of courses at other institutions. We plan to cover machine learning design, privacy architecture, fairness in visual recognition [and a bunch of other course topics listed here].
3) Make the role-plays streamlined for use outside of computer science courses, especially by civil society organizations looking to increase access to computing professions for underrepresented groups. We will spearhead this effort through a partnership with AI4All, training undergraduate volunteers to facilitate these role-plays with their near-peer mentees (including mobile play). (1167 characters)

Suitability (2500 characters or fewer)

Arvind Naryayan is Professor of Computer Science, and Incoming Director of our partner the Center for Information Technology Policy, the hub for all AI Ethics activities at Princeton. He has received multiple NSF awards in AI Ethics, published over 15 papers on the topic for a decade, and many more on the related topic of privacy. In addition, other areas of his AI Ethics research include dark patterns, online political ad targeting, and harms of recommender systems. These publications have collectively been cited over 8,000 times.

Professor Narayanan dedicates significant time to education outreach. He is a regular instructor for our partner AI4ALL, introducing AI to high schoolers, esp. female and minority students. He developed a Coursera with over half a million learners. He has taught multiple CS courses touching on ethics. In fact, he taught one of the first courses (at any university) on fairness and machine learning in fall 2017 (Barocas and Hardt taught on that topic concurrently).

His involvement in public scholarship on AI Ethics is extensive. He served on the inaugural steering committee of the FAccT conference, has delivered addresses on AI ethics at the AIES and NAACL, named lectures at MIT and Princeton, and a keynote at an Apple ML summit. He coauthored a textbook on fairness and ML (MIT press, forthcoming); the draft has been cited over a 1,000 times and used in a dozen courses. He is also coauthoring a trade book on AI snake oil, under contract with Princeton Press.

Steven Kelts teaches Ethics in the School of Public and International Affairs and the University Center for Human Values at Princeton. He began "Agile Ethics" with a seed grant from our Council on Science and Technology, developing role-plays about a campus mental health app, video generation, and more. He also runs the GradFutures cohort in A.I. Ethics, encouraging Ph.D. students in all fields to shift their research towards tech ethics, and helping others find industry positions in the space. He received a University award for this work.

His research is supported by a grant from Google to study "Ethics of Innovation," building upon his work on theories of ethics, innovation, and corporate structure.

Dr. Kelts's other focus is first-gen and low-income student success. For four years he worked on the curriculum for Princeton's nationally-recognized FGLI program, the Freshman Scholars Institute. In 2020, he banded together

with former students to found Kalos Academy, an all-volunteer non-profit mentoring FGLI students. He represents Kalos on the Association of Governing Boards' national Council for Student Success.

Concept Description (5000 characters or fewer)

Technologists can be caught flat-footed, failing to anticipate the social and moral implications of their work. How do we teach students cutting-edge technical skills while also raising their awareness of potential pitfalls before they happen?

Our approach, "Agile Ethics," brings together technical skills development through coursework with moral awareness through interactive role-plays. Over the next few years, we aim to integrate these role-plays into intro-level CS courses; run them for Social Science, Humanities and other STEM majors through our Council on Science and Technology; and use them to introduce underrepresented high school students to the design process through the non-profit AI4ALL.

This fall, Professor Arvind Naryanan will launch a course entitled "Ethics of Computing," aiming to show advanced CS students how to control the powerful tools they have at their disposal. As he has in his published work for more than 15 years, he will emphasize that technical skill must come with social consciousness and a healthy dose of humility. He will equip students with the technical and moral tools to assess what they have learned in their CS education, and think critically during their careers no matter how technology changes.

These techno-moral skills will be reinforced through "Agile Ethics"role-plays that simulate the development of products using the tech taught in class. Moral awareness begins with foreseeing the sometimes-hidden consequences of any action (Rest, 1986; and others). In the Agile Ethics simulations, students inhabit roles on an Agile team (engineers, project managers, etc.), while others take on roles in stakeholder groups (corporate board members, tech writers, end users, etc.) to experience ethical dilemmas as they emerge in real tech work. The students make technical and design choices, then discover through successive "sprints" that their choices may violate social norms.

The development of these simulations began a year ago, with insights from the literature on organizational psychology. It is rarely the case that employees of a company knowingly choose what's wrong; they are more often blinded from the consequences of their actions by financial incentives, workplace competition, etc. (See e.g., Max Bazerman and Ann Tenbrunsel, 2012.) The Agile Ethics role-plays mimic the stressful corporate environment. Initial studies of our pilot – performed by a team in our Psych Department – show a difference between respondents who are given a technomoral scenario that explicitly states ethical concerns and those given scenarios that don't. Those given morally-labeled cases may develop a false confidence (as measured by two perceived competence metrics) that they are equipped to make moral

choices in the future. But when shown morally-unlabeled cases, these same people are no more sensitive to potential dangers than any others.

Agile Ethics attempts to operationalize technomoral "humility" in these precise situations. Rather than using morally-labeled cases, it puts students in the position to experience how technical choices have unforeseen ethical consequences. Our Psych partners are now studying whether students who have gone through the role-plays will express more moderate confidence in their moral foresight, and understand better the need to learn from affected stakeholders, to gather deeper technical and social information, to resist organizational influences on their decision-making, and so on.

Our approach was developed in cooperation with responsible innovation professionals at a major tech company. Their goal has been to facilitate better decisions by actual engineering teams; ours is broader interdisciplinary engagement with ethics, including basic principles like harms, duties, virtues, justice, structural inequity, care, etc. By mirroring industry processes we also teach practical lessons about Agile, design thinking, and team-centered decision-making.

Our program raises awareness of the ecosystem in which tech is created by engaging directly with diverse stakeholders. Our pilot was co-designed with industry professionals who came to campus to share their experiences. Our students saw that their activity was built on real scenarios, encountered by professionals who looked like them. As we expand, we will partner with the NYC-based non-profit All Tech Is Human to engage a broader base of professionals from all backgrounds, and to host events with area universities to share best practices for Responsible Computing instruction. On our own campus, we'll also partner with our Tech Policy Clinic to engage policymakers in the role-plays.

Feasibility (1250 characters or fewer)

1) We currently use an effective (though simplified) web-based scoring system to award points to "engineers" and "stakeholders." While high scorers don't "win" anything tangible, the gamification reinforces the incentives in the character descriptions (like in a game of Clue). To serve our partner non-profit AI4ALL and other universities we will need new tech for remote game play on a mobile app. We are obtaining bids from organizations experienced in gamification.

2) We design each of our role-plays with industry professionals, aiming to be true-to-life. We will engage undergraduates in the process, to do interviews and additional research. With one summer to improve upon our pilot, coordination among many individuals will be a challenge. Professor Kelts will coordinate with All Tech Is Human to involve diverse NYC-based professionals, and mentor the student co-designers.

3) Like many programs teaching responsible computing, Agile Ethics saw barriers to buy-in. Students laser-focused on marketable technical skills don't always see the relevance of ethics. But having industry professionals speak about their experience helps bypass these barriers, showing students that they are simulating real (and important) professional choices.

Working Open (1250 characters or fewer)

The true measure of success for Agile Ethics will be its use in contexts beyond the classroom, to inspire budding programmers, future policymakers, and any concerned stakeholders. So we will make our role-plays and topics from the "Ethics of Computing" course freely available on the website of Agile Ethics: https://agile-ethics.princeton.edu/. Anyone can use them to teach tech design, and we will partner with AI4ALL to ensure they are used to empower disadvantaged students to take up careers in responsible computing.

By including undergraduates in the development of our role-plays, we will engage the Humanities and Social Sciences as well. To create a just tech future, we need them as much as we need responsible programmers. By putting students in contact with industry professionals through our partner All Tech Is Human we'll foster the interdisciplinary conversation. And as we expand the use of the role-plays across CS courses and to other disciplines (beyond the Policy course where they are already in use) that conversation will only grow. By partnering with our Council on Science and Technology, we're committing to their mission to "advance STEM literacy" through interdisciplinary programming and curricula.

Impact (2500 characters or fewer)

Exploring the Agile method and its pitfalls can foster the ethical growth of responsible technologists and improve the design processes they use, as well as inform regulators and the public about how to create a more just tech future.

Assessment of participants' "Moral Awareness" is integral to Agile Ethics. Moral Awareness, as defined by James Rest (1986) is a decision-making skill that requires information-seeking, social cognition (relating self to others) and situational understanding. In the context of tech work, this requires knowledge of and sensitivity to stakeholder needs, the strength to resist workplace financial and social pressures to maintain focus on others, and the humility to admit imperfect situational knowledge (i.e., that there can be dangers we fail to foresee). With our partners in the Concepts and Cognitions Lab we have designed a novel psychological metric to assess changes in Moral Awareness – including measurements of subjects' perceived competence to

address unknown circumstances, their perceived need for accurate information, and their sensitivity to others' concerns. Preliminary results have been promising; a complete study of the effects of our pilot on increasing Moral Awareness will be completed this summer.

From experience, we know that Agile Ethics has the potential to seriously impact the moral development of future technologists. Participant responses to our initial pilot with CST – especially from students who had done internships at tech firms – confirmed that it helped them to reflect upon their experience and prepare themselves socio-emotionally for future career challenges. When used in the context of Professor Narayanan's course (and future CS courses), these role-plays will clearly convey the message that technical skill is inseparable from ethical professional conduct and reflective team decision-making. The family of Agile development methods increases innovation and efficiency in tech corporations – we hope to inspire our students to improve the ethical potential of these processes as well.

Agile Ethics role-plays also inform future policy makers about the work of tech development, and the interventions that could help regulate the production of safer technologies. The pilot role-plays are currently in use in Professor Kelts's Policy course entitled "Tech/Ethics". They are already inspiring discussions about, e.g., Federal regulation of the design process akin to financial audit regulations, etc.

Field Survey (1250 characters or fewer)

Studies suggest the utility of role-plays for limited purposes: teaching engineering ethics codes (Herkert 2000; Loui 1999), responsible conduct of research principles (Antes et.al. 2009; Loui 2010), and socially-conscious design (Doorn & Kroesen 2013).

Joe Herkert at N.C. State is a pioneer in role-plays to teach engineering ethics. But he suggested that these work for "microethics" (rules, codes, theories) not "macroethics" (social effects) (2005). Recent studies are more hopeful about teaching social dimensions of engineering practice with role-plays (see Martin et.al. 2019). Michael Loui at U. Illinois has used extended simulations to successfully teach tech and society themes. (2009) Ahish Hingle, Aditya Johri et.al. at George Mason have done the same for computing students. (2021) And some Dutch scholars have improved concern for social effects by teaching design thinking. (Lloyd & van de Poel 2009)

Agile Ethics uses short-term role-plays incorporated into a technical course, utilizing design principles and workflow terminology. We cannot be satisfied with students professing they are aware of social effects (or of codes or principles), if they don't have practical knowledge of when and how to intervene in daily work.

Budget Overview (2500 characters or fewer)

Line Item:	Explanation:	Request:
App Design, Construction	Obtaining bids from Fascination Lab, Workverse and others	**$40,000**
All Tech Is Human	Co-sponsored event in NYC on effective teaching of responsible computing, and building the responsible tech university network. This partnership is vital to securing the involvement of a diverse group of industry professionals in our nearest city.	**$10,000**
Student Stipends	Funding for 3 undergraduate RA's to develop experiential learning materials	**$15,000.00**
Faculty Summer Salary	Summer funding for Kelts to direct RA work and oversee app design.	**$15,000.00**
Honoraria and Travel for Campus Visitors	Visits by tech industry professionals. All sessions exploring real use-cases from their companies.	**$7,000.00**
Miscellaneous Support for Co-Curricular Program	Refreshments, materials, etc. The co-curricular use of our role-plays by the Council on Science and Technology, and by AI4ALL, will require involvement of departmental admins	**$3,000.00**
University Overhead	Grant conditions state max 10%	**$10,000**
	TOTAL	**$100,000.00**