Ethical Issues in Computing and AI

Research Questions and Pedagogical Strategies

9:00 AM to 6:00 PM

June 2–3, 2022

Harvard University

LARGE PRINT AND SIMPLE TEXT PROGRAMME

June 2

9:00 AM	**Opening Remarks** Barbara Grosz and Alison Simmons (Harvard)
9:15 AM	**Keynote** Luciano Floridi (Oxford)
10:30 AM	**15-minute break**
10:45 AM	**The Role of Imaginative Variation in Ethics by Design** Stacy Doore (Colby) and Fernando Nascimento (Bowdoin)
11:30 AM	**Legitimacy of What? A Call for Democratic AI Design** Jonne Maas (TU Delft)
12:15 PM	**Lunch**
1:15 PM	**Click-Gap isn't Paternalist, Epistemic or Otherwise — And that's a good thing** J. L. A. Donohue (Harvard)

2:00 PM **Autobiographical Thinking as a Pedagogical Tool: Envisioning Oneself as a Protagonist of Ethics**

Omowumi Ogunyemi (Pan-Atlantic)

2:45 PM **15-minute break**

3:00 PM **Keynote**

C. Thi Nguyen (Utah)

4:15 PM **15-minute break**

4:30 PM **Greater Boston Area AI Ethics Lightning Round**

John Basl, Jeffrey Moriarty, Meica Magnani, William Cochran, William Griffith, J. L. A. Donohue, Jordon Kokot, Vance Ricks, Crystal Lee, Kay Mathiesen

5:30 PM **30-minute break**

6:00 PM **Outdoor reception**

June 3

9:00 AM	**Keynote** David Kaiser (MIT)
10:15 AM	**15-minute break**
10:30 AM	**Biased-by-Design: Why Algorithms are Necessarily Value-Laden** Phillip Kieval (Cambridge)
11:15 AM	**Examining Professional Social Responsibility Development among Computer Science Undergraduates Using the Generalized Professional Responsibility Assessment** Quintin Kreth (Georgia Tech)
12:00 PM	**Lunch**
1:00 PM	**Violence and Cyber-Violence** Kiran Bhardwaj (Andover)

1:45 PM **Educating for Moral Agency in an Ethics of Emerging Technologies Course**

William Cochran (Harvard)

2:30 PM **15-minute break**

2:45 PM **The Right to be an Exception in Data-Driven Decision-Making**

Sarah Cen (MIT) and
Manish Raghavan (Harvard)

3:30 PM **Avoiding Harms or Promoting the Good? Two Approaches to Embedded Ethics in Computer Science Education**

Avigail Ferdman (Technion)

4:15 PM 15-minute break

4:30 PM **Keynote**

Alison Simmons (Harvard), James Mickens (Harvard), and Kathy Pham (Harvard Kennedy School and Mozilla Foundation)

5:30 PM **Closing Remarks**

Keynote Speakers

Luciano Floridi is the Oxford Internet Institute's Professor of Philosophy and Ethics of Information at the University of Oxford, where he is also the Director of the Digital Ethics Lab, and Professorial Fellow of Exeter College. His research concerns primarily Information and Computer Ethics (aka Digital Ethics), the Philosophy of Information, and the Philosophy of Technology. He has published over 150 papers in these areas. He is also deeply engaged in civic and corporate policy development on emerging technologies with the UK, Germany, the European Commission, Cisco, Google, IBM, Microsoft, and Tencent.

C. Thi Nguyen is a former food writer, now Associate Professor of Philosophy at the University of Utah. He writes about trust, art, games, and communities. He is interested in the ways that our social structures and technologies shape how we think and what we value. He has published dozens of articles on these topics. His first book, Games: Agency as Art, was awarded the American Philosophical Association's 2021 Book Prize. He

has also written popular pieces for outlets such as Boston Review and The New York Times, and has appeared on podcasts including The Ezra Klein Show and Philosophy Talk.

David Kaiser is Germeshausen Professor of the History of Science, Professor of Physics, and Associate Dean for Social and Ethical Responsibilities of Computing (SERC) at MIT. His historical research focuses on the development of physics in the United States during the Cold War, looking at how the discipline has evolved at the intersection of politics, culture, and the changing shape of higher education. His physics research focuses on early-universe cosmology, working at the interface of particle physics and gravitation. He has also helped to design and conduct novel experiments to test the foundations of quantum theory.

Alison Simmons is Samuel H. Wolcott Professor of Philosophy at Harvard University. With Barbara Grosz, she is co-founder of the Embedded EthiCS programme, which develops ethics modules for computer science courses at Harvard. Her

research interests lie primarily at the intersection of philosophy and psychology. She works on questions about the nature of mind in general, the nature of sense perception in particular, and conceptions of the relation between mind and world as they have developed historically from the ancient through the medieval and early modern periods, and also as it is discussed today.

James Mickens is Gordon McKay Professor of Computer Science at the Harvard John A. Paulson School of Engineering and Applied Sciences. He studies distributed systems, specifically, how to make them faster, more robust, and more secure. Much of his work focuses on large-scale web services, and how to design principled system interfaces for those services. He has also given many talks offering his unique brand of wisdom to all who will listen. In his spare time, he enjoys life, liberty, and the pursuit of happiness, often (but not always) in that order, and usually (almost always) while listening to Black Sabbath.

Kathy Pham is a computer scientist, product leader, and founder with experience across the

private and public sectors, and a love for developing products, building and leading teams, data, healthcare, and weaving public service and advocacy into all aspects of life. She is a Senior Advisor for Responsible Computing at the Mozilla Foundation, a Senior Fellow at the Shorenstein Center at the Harvard Kennedy School of Government, Founder of Product and Society and the Ethical Tech Collective, and Deputy Chief Technology Officer of the Federal Trade Commission.

Abstracts

The Role of Imaginative Variation in Ethics by Design

Stacy Doore (Colby College) and
Fernando Nascimento (Bowdoin College)

We present the Computing Ethics Narratives website as a platform to foster the integration of imaginative variations in the conceptualization and design of new digital technologies. In the first part, we offer a summary of the theoretical framework that supports the role of creative imagination in ethical deliberation. In particular, we highlight the possibilities of fictional narratives as ethical laboratories that expand the diversity of scenarios considered in the deliberative process. In the second part, we present the website Computing Ethics Narratives as a tool to foster ethical sensibility in computer science students and related areas through fictional narratives.

The project can be accessed at:
https://www.computingnarratives.com/

Legitimacy of What? A Call for Democratic AI Design

Jonne Maas (Delft University of Technology)

The legitimacy of AI decision-support systems raises several concerns. Especially in the field of public decision-making, the consequences of the use of opaque machine learning systems have been under critical investigation as these systems jeopardize democratic rights like transparency and contestability. To question how such AI systems affect democratic rights is worthwhile, as legitimacy is concerned with justified exercises of public (i.e. the state's) power. However, the focus on the system itself (and its consequences for democratic rights) overshadows a distinct type of legitimate decision-making, namely the legitimacy of the design decisions underlying the development process of an AI system. Based on two contextual case studies, one in the legal (public) domain and one in the medical (private) domain, we argue that the legitimateness of an AI system predominantly depends on legitimate design decisions, which we claim should be rooted in democratic ideals. First, following political

philosophers like Rawls and Cohen, democratic procedures contribute to overall fairness. Second, following political philosophers like Pettit, we conceive of design decisions as exercises of power that (can) have an effect on the public realm, therefore requiring these decisions to be grounded in democratic control.

Click-Gap Isn't Paternalist, Epistemic or Otherwise—and That's a Good Thing

Jenna Donohue (Harvard University)

Scholars have argued that some recent interventions made by social media companies, such as Facebook's "Click-Gap", are instances of epistemic paternalism because they are undertaken with the goal of improving the epistemic status of the users. I think arguments of this sort face important problems, problems we should take seriously as we begin to recognize these platforms require improvement and regulation. While interventions like Click-Gap may make social media platforms slightly less bad, they do not represent nearly enough change to tackle

the scale of the problems that these platforms have introduced into our lives, our communities, our democracy. Further—calling interventions of this kind justified epistemic paternalism troubles me. It gives the impression that the interventions are prima facie unjustified because paternalistic, making it seem as though (1) the companies' doing nothing would be prima facie justified (because not paternalistic) and (2) we should focus our scholarly attention on defense of such interventions. Both (1) and (2) are problematic. I think it isn't true that the companies' doing nothing would be prima facie justified, and I think our scholarly attention ought to be focused on if and how we might resolve some of the many problems these platforms have brought into our lives—and what their responsibilities are for doing so. In this paper I argue that Click-Gap is not paternalistic, contrary to the conclusion of others. Then, I will say a bit about why this matters: though there may be circumstances in which paternalistic behavior is justified, it isn't in cases involving the sort of relationship that Facebook has with its users. Finally, I suggest some implications for the

relationship between institutions and paternalism more generally.

Autobiographical Thinking as a Pedagogical Tool: Envisioning Oneself as a Protagonist of Ethics

Omowumi Ogunyemi (Pan-Atlantic University)

In an educational institution with people of different cultural backgrounds and religious beliefs, one often finds students who have different references for ethical decision-making. It is not unusual to find those reluctant to discuss ethics if they feel that their views or tenets for practices may be shaken by novelty of approaches and views to ethical work. Moreover, students may not show much interest in reading the articles and texts in the humanities including ethics when they are required subjects, especially when they do not see its relevance for their lives. In addition to the challenges caused by diversity, it is common to find people complaining about the problems of the society without thinking of how they can contribute to solving it. How then can one engage science

students in reading philosophical texts, exploring topics in ethics with the hope of demonstrating its practical implications for life? How can one help them see its rich contributions to peaceful human coexistence and human flourishing, irrespective of one's backgrounds and beliefs?

This paper proposes storytelling and autobiographical thinking as means of integrating the elements ethics into students' curriculum, by building up stories in line with selected issues in the humanities. The pedagogical tool has two aims, firstly to introduce the students to ethical issues in AI and computer science and secondly to help them envision themselves as agents of change, in a society where it is more common to wait for the government to solve every need or to blame the government for many daily unpleasant experiences, including those that can be solved by ordinary citizens. The paper will be a basis for developing a curriculum that prepares the future leaders in the world of science. It is hoped that each student who uses this curriculum will identify areas in which they can use their knowledge of computing and information sciences to make a

changes that promote human flourishing for their selves and for others in the society.

Biased-by-Design: Why Algorithms are Necessarily Value-Laden

Phillip Kieval (University of Cambridge)

Algorithmic decision-making systems applied in social contexts drape value-laden solutions in an illusory veil of objectivity. I argue that these systems are necessarily value-laden and that this follows from the need to construct a quantifiable objective function. Many researchers have convincingly argued that machine learning systems learn to replicate and amplify pre-existing biases of moral import found in training data. But these arguments permit a strategic retreat for those who nevertheless maintain that algorithms themselves are value-neutral. Proponents of the value-neutrality of algorithms argue that while the existence of algorithmic bias is undeniable such bias is merely the product of bad data curation practices. On such a view, eliminating biased data would obliterate any values embedded in

algorithmic decision-making. This position can be neatly summarized by the slogan “Algorithms aren’t biased, data is biased.” However, this attitude towards algorithms is misguided. Training machine learning algorithms involves optimization, which requires either minimizing an error function or maximizing an objective function by iteratively adjusting a model’s parameters. The objective function represents the quality of the solution found by the algorithm as a single real number. Training an algorithm thus aggregates countless indicators of predictive success into a single, automatically generated, weighted index. But deciding to operationalize a particular goal in this way is itself a value-laden choice. This is because many qualities we want to predict are qualitative concepts with multifaceted meanings. Such concepts like “health” or “job-applicant-quality” lack sharp boundaries and admit plural and context-dependent meanings. Collapsing concepts into a quantifiable ratio scale of predictive success flattens out their quality dimensions. This process is often underdetermined and arbitrary, but convenient for enterprises that rely on precise and unambiguous predictions. Hence, the very choice

to use an algorithm in the first place reflects the values and priorities of particular stakeholders.

Examining Professional Social Responsibility Development among Computer Science Undergraduates Using the Generalized Professional Responsibility Assessment

Quintin Kreth (Georgia Institute of Technology)

For many years, scholars and public figures have called for improved ethics education in computer science degree programs to better address emerging societal issues in technology. Despite these calls, there is limited empirical evidence regarding professional ethics attitudes among computing students and the factors that influence their development. This gap is notable relative to studies of students in other science and engineering disciplines, where empirical studies of professional ethics development have a long history. This knowledge gap makes it difficult to design and evaluate ethics education interventions in computer science.

In this presentation, we present a survey instrument for measuring the development of professional social responsibility among computer science students, the Generalized Professional Responsibility Assessment (GPRA), developed as part of an NSF-supported research project (Award #1635554). Based on the Professional Social Responsibility Development Model (PSRDM) and its associated, validated survey instrument developed by Canney and Bielefeldt (2015), the GPRA measures “social responsibility attitudes” (SRAs) across eight “dimensions” and three summative “realms.” We administered the GPRA survey to a sample of 982 students graduating from an undergraduate program (including 184 computing majors) at a large engineering institution. Using these data, we examine SRAs cross-sectionally among computing students and other students, to identify variation between academic disciplines and trends within computing. We find that computing students have lower SRAs than their peers in other STEM disciplines. Further, the data indicate that male computing students have lower SRAs than their female peers. We discuss the advantages and disadvantages of the

GPRA and the PSRDM in the context of current computing ethics assessment options. We also discuss the value of quantitative and mixed-methods work in studies of computing ethics, options for improving future surveys, and opportunities for further research.

Violence and Cyber-Violence

Kiran Bhardwaj (Phillips Academy Andover)

Christopher Finlay's (2018) "Just War, Cyber War, and the Concept of Violence" argues that some kinds of cyberattacks are morally equivalent to armed kinetic attack (358). As a result, these cyberattacks—'violent cyberattacks'—can be responded to with kinetic violence (and vice versa) (Finlay 374). I argue that Finlay's account of Violence—the Double-Intent Theory—is both too restrictive and too wide. I argue that it improperly includes harms to property as violence, and that Finlay's dismissal of structural violence is also incorrect. Instead, I argue for a view that violence is best understood as dominating harms to persons (as the locus of actions).

Educating for Moral Agency in an Ethics of Emerging technologies course

William Cochran (Harvard University)

The rapid pace of technological change often outstrips the ability of legislators and regulators to establish proper guardrails. A solution is for those who develop and use emerging technologies to develop themselves as moral agents. This presentation describes a course taught at Wake Forest University in Fall 2020 that sought to meet this need. It highlights aspects of the course designed to help students transform from learning about the ethics of emerging technologies to being leaders in the emergence of ethical technologies. It then shares the results of a mixed-methods study that used a pre-post design to examine the course's effectiveness in developing students' moral dispositions and character traits. The findings, plus students' comments on course evaluations, suggest that the course design indeed supported students' development as moral agents.

At the beginning of the course, students were introduced to three different ethical theories — utilitarianism, deontology, and virtue ethics—and discussed how each acted as a lens that illuminated different ethical concerns. Shannon Vallor's Technology and the Virtues was then used as an anchor text for the remainder of the course. Students completed an assignment called a Technomoral Virtue Field Journal, which prompted students to design a plan to cultivate one of Vallor's 'technomoral virtues' in themselves and reflect on their experience. The course's major assignment was a term paper comprised of these steps: (1) describe an emerging technology, (2) elicit the major ethical problems that it could create if left unchecked, (3) construct a code of ethics to address these problems, (4) discuss the code's underlying values, (5) respond to a potential objection. Deadlines were scaffolded throughout the semester, culminating in student presentations.

The Right to be an Exception in Data-Driven Decision-Making

Sarah Cen (Massachusetts Institute of Technology) and Manish Raghavan (Harvard University)

Data-driven assessments estimate a target—such as the likelihood an individual will recidivate or commit welfare fraud—by pattern matching against historical data. There are, however, limitations to pattern matching. Even algorithms that boast near-perfect performance on average can produce assessments that are inappropriate for specific individuals. From the assessment's point of view, these individuals are exceptions, and in some contexts, failing to recognize exceptions can lead to decisions that inflict irreparable harm on individuals through no fault of their own. In this Article, we study how overlooking exceptions can yield undesirable outcomes and how this observation already motivates notions in the law—such as dignity and the right to individualized sentencing—as well as research areas in computer science—such as causal inference and robust optimization. Although the belief that exceptions

matter to high-stakes decisions is not new, the absence of a legal framework that acknowledges the unique challenges around exceptions in data-driven contexts has left a large accountability gap in the governance of data-driven decisions. To close this gap, this Article proposes that individuals have the right to be an exception in data-driven decision-making. The right requires that, when a decision can inflict harm on an individual, the decision maker must consider the level of uncertainty that accompanies a data-driven assessment and, in particular, whether it is appropriately individualized. The greater the risk of harm, the more serious the consideration. In this Article, we unpack the right to be an exception in detail, examining how it necessitates that uncertainty be meaningfully incorporated into data-driven decisions, affects the legitimacy of and trust in algorithms, and rebalances the burden of proof between decision makers and subjects. We conclude by discussing ex ante and ex post legal measures and surveying related areas in algorithm design.

Avoiding Harms or Promoting the Good? Two Approaches to Embedded Ethics in Computer Science Education

Avigail Ferdman (Technion—Israel Institute of Technology)

Amid the growing interest in ethics integration into computer science education, ‘Embedded Ethics’ is emerging as an important pedagogy. Embedded ethics integrates philosophers directly into computer science courses, to teach students how to think through the ethical and social implications of their work. This paper offers one of the first systematic reflections on embedded ethics. It presents two approaches to doing embedded ethics. The first approach is ‘avoiding harms’. On this approach, the emphasis is on teaching students to avoid designing technologies that are harmful, create bias or entrench systemic inequalities. The second approach is ‘promoting the good’. On this approach, the emphasis is on teaching students to create technologies that enable people to lead good lives. The two approaches are developed using a neo-Aristotelian heuristic. Humans do well when they develop their

human capacities, to know, to create, to be moral and the capacity to exercise the will. Technology shapes the environments in which persons develop and exercise their human capacities. Some technologies create environments that limit persons' ability to develop and exercise their capacities. For example, addictive-by-design technologies actively work against a person's capacity to exercise the will; media platforms that spread misinformation limit one's capacity to know. Alternatively, some technologies create environments that encourage the development and exercise of human capacities. For example, technologies that use open source enable users to co-design and in effect encourage the capacity to create. The 'avoiding harms' approach will train student to avoid designing technologies that limit persons' ability to develop and exercise capacities. The 'promoting the good' approach will train students to design technologies that encourage the development and exercise of human capacities. While each approach emphasizes different normative concerns, they can be complementary.

Organizers: Trystan Goetze, Kevin Mills

Our Thanks to: Jill Susarrey, Alison Simmons, James Mickens, Kathy Pham, David Kaiser, C. Thi Nguyen, Luciano Floridi, Jeff Behrends, Barbara Grosz, Julie Shah, Jenna Donohue, William Cochran, Ashley Bens, Matt Kopec, John Basl

Funders:

- The Faculty of Arts and Sciences at Harvard University
- The Harvard John A. Paulson School of Engineering and Applied Sciences
- The Social and Ethical Responsibilities of Computing initiative of the Schwarzman College of Computing at the Massachusetts Institute of Technology
- The Responsible Computer Science Challenge, which is funded by:
  - The Omidyar Network
  - The Mozilla Foundation
  - Schmidt Futures
  - Craig Newmark Philanthropies

Very Special Thanks to: Songyee Yoon