2023

Ethics4eu: Designing New Curricula For Computer Science Ethics Education: Case Studies For Ai Ethics

Dympna O'SULLIVAN
Technological University Dublin, Ireland, dympna.osullivan@tudublin.ie

J Paul GIBSON
Telecom Sud Paris, paul.gibson@telecom-sudparis.eu

Andrea CURLEY
Technological University Dublin, Ireland, Andrea.F.Curley@TUDublin.ie

See next page for additional authors

ETHICS4EU: DESIGNING NEW CURRICULA FOR COMPUTER SCIENCE ETHICS EDUCATION: CASE STUDIES FOR AI ETHICS

D O'Sullivan [1]
School of Computer Science, Technological University Dublin
Dublin, Ireland
ORCID: 0000-0003-2841-9738

JP Gibson
Institut Polytechnique de Paris, TELECOM SudParis
Évry, France
ORCID: 0000-0003-0474-0666

A Curley
School of Computer Science, Technological University Dublin
Dublin, Ireland
ORCID: 0000-0001-9412-8512

A Becevel
School of Computer Science, Technological University Dublin
Dublin, Ireland
ORCID: 0000-0001-7704-8975

E Murphy
School of Computer Science, Technological University Dublin
Dublin, Ireland
ORCID: 0000-0001-6738-306

D Gordon
School of Computer Science, Technological University Dublin
Dublin, Ireland
ORCID: 0000-0002-3875-4065

[1] *Corresponding Author (All in Arial, 10 pt, single space)*

D O'Sullivan

dympna.osullivan@tudublin.ie

***Conference Key Areas**: Education about and education with Artificial Intelligence, Embedding Sustainability and Ethics in the Curriculum*
***Keywords**: AI Ethics, Case Studies, Curriculum Development*

ABSTRACT

The computing ethics landscape is changing rapidly, as new technologies become more complex and pervasive, and people choose to interact with them in new and distinct ways. The resultant interactions are more novel and less easy to categorise using traditional ethical frameworks. It is important that developers of these technologies do not live in an ethical vacuum, that they think about the consequences of their creations, and take measures to prevent others being harmed by their work. To equip developers to rise to this challenge and create a positive future for the use of technology, it important that ethics becomes a central element of computer science education. To this end, the Ethics4EU project has developed curricula on a wide range of topics including privacy and agency of personal information, digital literacy, data governance and accountability, surveillance applications, algorithmic decision and automating human intelligence for robotics and autonomous vehicles. Crucially the content examines computing ethics, not only in terms of hardware and software, but how systems, people, organisations and society interact with technology.In this paper, we present our interdisciplinary approach to developing educational content for AI Ethics. This includes accessible teaching materials, in-class activities, sample assessments, practical guidelines and instructor guides. We discuss findings of an evaluation of the developed content with undergraduate computer science students.

---

1 INTRODUCTION

1.1 Background

Computers and technological applications are now central to many aspects of life and society, from industry and commerce, government, research, education, medicine, communication to entertainment systems. These technologies have wide ranging impacts on society, and despite the many ways new technologies have improved life, they cannot be regarded as unambiguously beneficial or even value-neutral.There is a sense that some technology development and innovation is happening at a more rapid pace than the relevant ethical and moral debates.

The history of computing ethics (or computer ethics) goes hand-in-hand with the history of computers themselves; since the early days of the development of digital computers, pioneering computer scientists, such as Turing, Wiener and Weizenbaum, spoke of the ethical challenges inherent in computer technology [1], but it was not until 1985 that computing ethics began to emerge as a separate field. This was the year that two seminal publications were produced, Deborah Johnson's book Computer Ethics [2] and James Moor's paper, "What Is Computer Ethics?" [3]. Deborah Johnson's *Computer Ethics,* was the first major book to concentrate on the ethical obligations of computer professionals, and thoughtfully identifies those ethical

issues that are unique to computers, as opposed to business ethics or legal ethics. In James Moor's paper [3], he defined computer ethics as "the analysis of the nature and social impact of computer technology and the corresponding formulation and justification of policies for the ethical use of such technology", and argues that computer technology makes it possible for people to do a vast number of things that it was not possible to do before and since no one could do them before, the question may never have arisen as to whether one ought to do them.
In the 1990s, and the concept of "value-sensitive computer design" emerged, based on the insight that potential computing ethics problems can be avoided, while new technology is under development, by anticipating possible harm to human values and designing new technology from the very beginning in ways that prevent such harm [4]. Others including Donald Gotterbarn [5], theorised that computing ethics should be seen as a professional code of conduct devoted to the development and advancement of standards of good practice for computing professionals. This resulted in the development of a number of codes of ethics and codes of conduct for computing professionals, for example the ACM "Guidelines for Professional Conduct".
In 1996 the "Górniak Hypothesis" predicted that a global ethic theory would emerge over time because of the global nature of the internet. Developments since then appear to be confirming Górniak's hypothesis and have resulted in the metaphysical information ethics theory of Luciano Floridi [6]. These new theories make explicit the social and global change created by new technologies and call for an intercultural debate on computing ethics in order to critically discuss their impact on society.

1.2 Ethics4EU

The Ethics4EU project [7], is exploring issues around the teaching of ethics in computer science curricula. To understand gaps in the prorvision of ethics education and how to address then, the project undertook a pan-European survey of attitudes of computer science faculty towards teaching computing ethics [8]. The survey was completed by faculty at 61 universities across 23 different European countries. This found that 36% of respondents (or 22 universities) do not teach any computer ethics, citing either a lack of available time or a lack of expertise as being the key reasons as to why they don't teach this topic. When institutions do teach Computer Ethics, they tend to devote a relatively small number of hours to teaching Computer Ethics on their Computer Science or related programmes - 67% of institutions surveyed teach 10 hours or less per semester. Our survey also revealed that computer ethics is often taught as a standalone subject. This is in spite of evidence that infusing computer ethics in Computer Science curricula gives students a better understanding of the ethical impacts and possible harmful effects of the technologies they implement.

Research has consistently shown that ethics is an important missing element in computer science education unlike all other science disciplines [9]. Furthermore our survey results show that there is a lack of staff availability and expertise to teaching computing ethics [8]. Thus one of the key objectives of this project is to develop a 'train the trainer' range of teaching content and instructor guides to facilitate computer science faculty in the instruction of computing ethics. In this paper we present and evaluate educational content that was developed as part of the project, specifically lessons that focus on ethics related to computer programming errors.

The content is designed to serve as a way to improve computer science students' ability at consequence scanning – a way to consider the potential consequences of new software on people, communities and the planet [10].

2 METHODOLOGY

2.1 Case Studies

In this work, we describe the development and evaluation of educational content that addressed AI (Artifical Intelligence) ethics. A case study method was chosen as these are designed to explore real-world phenomena and they focus on interpreting events and exploring the impact of the case study on the broader society, including ethical issues [11]. Four case studies were carefully selected and developed by interdisciplinary teams of Computer Scientists and Ethicists, which focused on a number of programming-related ethical scenarios. The four case studies are briefly described below. The full case studies plus the in-class activities used to deliver the case studies are available at (Ethics4EU website [7]):

2.1.1 Irish State Examinations 2020

As in many countries, the Covid-19 pandemic had a profound effect on Irish state examinations in 2020. Due to a national stay-at-home rule, state examinations were cancelled and replaced with an algorithmic estimated-grading system. A student's grade in each subject was estimated based on their expected performance combined with their School's statistical profile of achievement. A national standardisation process was applied to ensure a consistent standard in the estimated-grade process across the country. A subsequent review of the standardization process revleaed that the algorithm produced a disproportionately negative outcome for high performing students from historically low performing schools. As soon as the errors were detected, the affected students were identified and corrections made. However, delays in making these corrections meant that some students had not received correct offers for university places and had to wait to commence their third-level study in the following academic year.

2.1.2 Search Engine Bias

The Google auto-complete algorithm looks for common queries that match what a user starts to enter into the search box but also considers the lauguage of the query, the location a query is coming from, trending interest in a query and the user's past searches. Google's rationale for offering auto-completion is to provide a more personalised search experience, however there are many recorded instances where autocompletion makes poor or even problematic suggestions that have prioritized sites with extremist biases.

2.1.3 Judicial Sentencing Software

Some courts of law in the United States of America are employing commercial software systems to assist the judiciary in sentencing criminal defendants. A ProPublica analysis of one of these sentencing systems, the COMPASS system found evidence of racial bias when making a sentencing recommendation [12]. The team found that “blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend,” whereas COMPAS “makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower-risk but go on to commit other crimes ». They also found that only 20 percent of people predicted to commit violent crimes actually went on to do so.

2.1.4 Autonomous Vehicles

In recent years, the automobile industry has seen some car manufacturers incorporating self-driving as an available feature. This feature enables the car to autonomously navigate between two geographical points without any, or minimal, intervention by the driver. The car uses an array of sensors to capture data in its environment, which is input to software controlling the car’s mobility and navigation. The software developed to enable autonomous self-driving must be capable of responding to the threat of a potential or imminent accident. When implementing these algorithms, programmers need to be cognisant of parameters that might include legal, moral, cultural, ethical and geographical factors.

2.2 Evaluating the Case Studies

We wanted to understand if computer ethics case studies highlight the importance of ethics for computer professionals and whether delivering the computer ethics case studies in a constructivistic manner help students see the case from multiple perspectives. One case study per week was delivered (over 4 weeks) as part of a first-year programming module between the 15th of April 2021 and the 30th of April 2021. The group composed of 175 first year computer science students at Technological University of Dublin, Ireland. The content was delivered using a virtual classroom for the main lessons, with breakout rooms for the students to discuss the ethical issues from each session in smaller groups, and Padlet (https://padlet.com/) was used as an idea sharing space where participants could highlight their key take-aways from each lesson. The lecturers recorded their reflections about the classes on a weekly basis in diaries. After the lessons were completed, the students were invited to participate in a survey to collect their feedback on the process.

3 RESULTS

A student survey was deployed using Microsoft Forms between the 30th of April 2021 and 3rd May 2021, and a total 25 students responded to the survey giving a

response rate of 14.3%. The students were given the following key instructions: (i) the survey is voluntary, (ii) all submissions do not record the students' names, and (iii) the results will be published as part of the broader discussion on these issues. The survey had seven questions (two closed-ended and five open-ended) and was developed based on a combination of the Learning Object Review Instrument (LORI) [13] and some examplars from Oppenheim's book on questionnaire design [14].

The findings were very instructive, including the fact that the majority of students (23 out of 25) rated the content as "Very Interesting". Specifically the students commented that it was interesting to explore how much of an impact the systems that they develop could have on other people's lives (as one commented "*how the programs … can potentially change someone's life for worse or better*"), particularly the case studies highlighted the potential dangers when computer programs are written in a rush or don't observe good programming practices.

The students also noted how the structure of these lessons differed from their typical classes, as they had more time to interact with their classmates, and hear different perspectives on a particular topic (one student commented that "*the case studies were different from regular lectures and felt like a fun TED talk, with additional audience engagement*"). This theme was further expanded upon in answers to different questions, including for example "*The group discussion helped to show different perspectives from my classmates*." and "*discussion about* [the cases] *and sharing of ideas helped me see things newly*". Interestingly, many students changed their minds on these topics based on these interactions, for example, "*I used to think technology was the answer to everything but after talking to others about* [it], *that really changed my perspective on it*". Also many commented on their expanded appreciation of ethics in programming, for example, "*I hadn't considered that there were ethics to think about in programming but having talked to people I realized now ethics can be applied everywhere*".

Another striking theme that emerged was how the case studies prompted the students to consider other people in the design of algorithms, for example, one noted that "*We all have different views on what ethics is so it is important that people from different backgrounds are always included when developing an algorithm*". Specific groups were mentioned as being especially important in this expanded perspectives, such as "*minorities and people with disabilities*" as well as "*the perspectives of the most vulnerable*".

The lecturers' diaries provided additional insight into the case study approach, they commented that the topics chosen were successful because they were highly relevant to the students and very tangible examples of the challenges inherent in software development. They also commented on the fact that it is difficult to find

good case studies that successfully balances the algorithmic aspects of the case with the ethical aspects, for example, in terms of the first case study on the Irish State Examinations, all of the participants were very much engaged in how a small algorithmic error can impact thousands of people, but they felt there wasn't significant ethical nuance to that particular case (when compared to the others). They felt the cases that involved more sophisticated Artificial Intelligence (the Judicial Sentencing Software and the Autonomous Vehicles Cases), were the ones that best balanced the ethical and technological considerations, and, in particular, the Autonomous Vehicles case provoked the most debate and controversy. The lecturers enjoyed this constructivistic approach to teaching, and where they gave the students time to work with each other, and truly reflect on novel and interesting topics.

4 ACKNOWLEDGMENTS

The authors of this paper and the participants of the Ethics4EU project gratefully acknowledge the support of the Erasmus+ programme of the European Union. The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

[1] Bynum, TW. 2001 "Computer ethics: Its birth and its future." *Ethics and Information Technology*, 3(2), 109-112.

[2] Johnson, D. 1985. "Computer ethics." Englewood Cliffs (NJ).

[3] Moor, JH. 1985. "What is computer ethics?" *Metaphilosophy*, 16(4), 266–275.

[4] Flanagan, M, Howe, DC and Nissenbaum, H. 2008. "Embodying values in technology: Theory and practice." *Information Technology and Moral Philosophy*, 322.

[5] Gotterbarn, D, Wolf, MJ, Flick, C and Miller, K. 2018. Thinking Professionally, the continual evolution of interest in computing ethics. *ACM Inroads*, 9(2), 10-12.

[6] Floridi, L. 2014. The Fourth Revolution: How the Infosphere is Reshaping Human Reality, OUP, Oxford.

[7] The Ethics4EU project: http://ethics4eu.eu

[8] Stavrakakis, I, Gordon, D, Tierney, B, Becevel, A, Dodig-Crnkovic, G, Dobrin, R, Schiaffonati, V, Pereira, C, Tikhonenko, S, Gibson, JP, Maag, S, Agresta, F, Curley, A, Collins, M, and O'Sullivan, D. 2021 "The Teaching of Computer Ethics on Computer Science and Related Degree Programmes: A European Survey", *International Journal of Ethics Education*, 6(2).

[9] Spradling, C, Soh, LK, and Ansorge, C. 2008. Ethics Training and Decision-Making: Do Computer Science Programs Need Help? Paper

presented at the *39th SIGCSE Technical Symposium on Computer Science Education, Portland OR USA*, pp. 153-157.

[10] Doteveryone, Consequence Scanning, An Agile Practice for Responsible Innovators, https://www.doteveryone.org.uk/project/consequence-scanning/, last accessed 2021/06/06

[11] Martin DA, Conlon E and Bowe B. 2018. A Constructivist Approach to the use of Case Studies in teaching Engineering Ethics. In: Auer M., Guralnick D., Simonics I.(eds) *Teaching and Learning in a Digital World. ICL. Advances in Intelligent Systems and Computing*, vol 715. Springer.

[12] Dressel, J and Hany F. 2021. “The Dangers of Risk Prediction in the Criminal Justice System.” *MIT Case Studies in Social and Ethical Responsibilities of Computing*, no. Winter 2021.