S. Draheim, J. Sudeikat (2023) ETHICS INSIDE? NOTES ON A BACHELOR CORE ELECTIVE COURSE IN COMPUTER SCIENCE, ICERI2023 Proceedings, pp. 8062-8067.

ETHICS INSIDE? NOTES ON A BACHELOR CORE ELECTIVE COURSE IN COMPUTER SCIENCE

S. Draheim, J. Sudeikat

Hamburg University of Applied Sciences (GERMANY)

Abstract

Our social and individual life in the 21st century is comprehensively shaped by accelerated technology and scientific development, in the world of work as well as in health, safety, mobility, communication and much more. As a result, previous social traditions, social practices and moral norms of living together are undergoing a process of fundamental change. We have concluded that it is necessary for prospective Computer Scientists to reflect on these questions and to develop an awareness of the scope of their own agency. To this end, it is very helpful for students to engage with debates, texts and questions on the topic of "machine ethics" from social sciences and humanities disciplines. Here, we present the agenda, the essential contents and first lessons learned from an elective course in the bachelor curriculum of the Computer Science programme at the Hamburg University of Applied Sciences (HAW Hamburg), which we held twice in the last semesters. In particular, we motivate addressing this interdisciplinary topic in a Computer Science degree programme. We outline the challenges of enabling interdisciplinary reasoning. Our practices, based on backgrounds in both social sciences and computer science research are outlined to motivate comparable efforts.

Keywords: machine ethics, computer science, social sciences, autonomous agents, adaptive systems

1 INTRODUCTION

The rapid technical progress in the development of artificial intelligence and autonomous systems presents ethical challenges [15]. Therefore, it is imperative to incorporate ethical education into computer science curricula to equip future computer scientists with the capability to recognize and analyze ethical issues in development projects. Assessing the impact of technical systems requires computer scientists to engage in ethical debates and argumentation, relying not only on factual and technical knowledge. This paper presents the outline and initial experiences of an elective course on machine ethics for computer scientists [1]. "[…] machine ethics is concerned with ensuring that the behavior of machines toward human users, and perhaps other machines as well, is ethically acceptable." [2, p.15].

The interdisciplinary character of the course subject is demonstrated by an iterative, alternating sequence of lectures and group discussions. Alongside this, inverted classroom sessions are also incorporated. These subsequent exercises are designed to engage students and move them from knowledge acquisition to critical analysis. Throughout this course, we explore the ethical significance of machines that are being designed and conceptualised with enhanced autonomy and intelligence. Specifically, we will examine the concept of a moral machine, its range of potential actions and its construction. We will draw upon current research and application in the field of machine ethics, situated at the intersection of computer science, philosophy and robotics. The course relates to ethical reasoning in the development of cyber-physical systems. This active research field delves into creating coherent systems of physical and logical components that demand autonomous and adaptive components, hence ethical quandaries arise. Additionally, we examine the initial applications where machine ethics is being utilised.

This paper is structured as follows. In the next section, related work is discussed. In the subsequent section, we present the course concept, i.e., we outline the learning goals, the structure of educational activities involved. This also concludes an analysis of the topics, which students have chosen for their paper assignments. Finally, we conclude and give prospects for future enhancements.

2 RELATED WORK

As Miller noted as early as 1988 “[…] technical issues are best understood (and most effectively taught) in their social context, and the societal aspects of computing are best understood in the context of the underlying technical detail.” [3, p.37]. This obviously also applies to ethical considerations, because since the 1980s, the topic of ethics for computer science students has been negotiated again and again, which is reflected in many publications on conceptual considerations, methodological or didactic approaches and some effectiveness studies on different forms of ethics with reference to computer science. [4; 5]. As autonomous systems and Artificial Intelligence related topics have become increasingly powerful and practical over the last five years, it is not surprising that concepts and publications specializing in machine ethics are also becoming increasingly visible in academia and in the context of teaching [6; 1].

Many seminar approaches focus on reflections on the future professional role. At the same time, it is crucial to perform ethical dilemma scenarios as realistically as possible. It has been found that a key challenge for ethics education is to engage students in recognizing and analyzing ethical challenges e.g., via discussing fictional scenarios [12]. While a variety of ethical codes and principles are available [8], handling ethical problems and considering conflicting principles requires individual analysis and debate in peer groups [8]. Approaches that embed ethical considerations, concepts and guidelines directly into basic subjects such as "Introduction to Computer Science" or "Programming" are particularly promising in this regard [9].

Despite ongoing research, studies consistently show that purely elective courses on ethics are insufficient. Additionally, a lively inter- or transdisciplinary cooperation between ethics and computer science is crucial for a successful curriculum [11].

Instead, it is necessary to experience the practical consequences of ethical reflection firsthand within the course [10]. The practical relevance of ethical reasoning in computer science cannot be disputed. Especially when developing Cyber-physical Systems and autonomous machines [12], it is imperative to consider that these systems become integrated or embedded within physical contexts and can subsequently interfere with and impact human stakeholders directly or indirectly. Additionally, ethical reasoning is significant concerning related areas of computational systems development that utilize algorithms for decision-making. This is exemplified by borderline paper topics as discussed in Section 3.

The course concept presented here centers on exploring interdisciplinary collaboration between computer science and social sciences. Our main objective was to provide clear and concise interlinking of ethical concepts with their informatic application references. We opted for a low-threshold seminar format with an elective course, focusing on text-based discussion and reflection work. Nevertheless, we intended from the outset to progress the module into a more hands-on and implementation-focused project-based course. This will be available for the first time during the upcoming summer semester.

3 THE COURSE CONCEPT

The seminar concept presented here was informed by previous experience with ethics content offered by one of the lecturers in an elective and as micro-lectures integrated into compulsory courses. However, collegial agreement was reached beforehand on the learning objectives we aimed to combine with specific ethics content, thus facilitating students' interdisciplinary competency acquisition.

We see four areas of competences and skills that are facilitated for preparing students (cf. Figure 1). First, a fundamental knowledge of *ethical theories and machine ethics* is required. To our experience, many students have limited prior knowledge and are particularly interested in this area. Evaluation of technical scenarios and example systems requires an understanding of technical capabilities and limitations for self-adaptive computing systems. Since students in different stages of their studies should be eligible to participate, a minimal technical foundation of intelligent adaptive Systems must be given as well. Engaging students in these areas requires group discussions on concrete example problems, thus participants engage in *ethical reasoning and argumentation*. Finally, self-sufficient and independent analysis requires proficiency in *literature research & scientific writing*, enabling students to acquire and express problem settings and varying viewpoints by themselves.

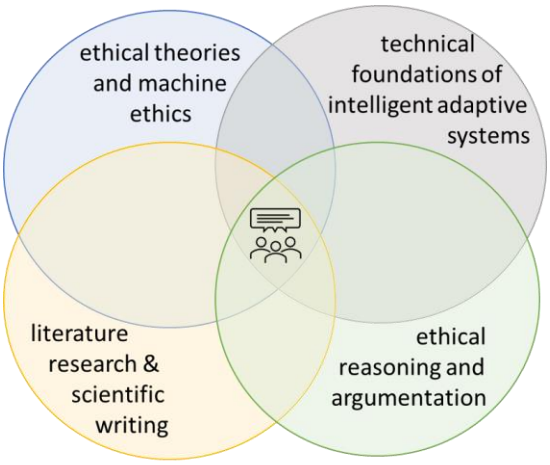

Figure 1. Required competences

The learning goals of the course derived from the required skills with respect to the working context of future computer science graduates. First, basic knowledge of fundamental ethical theories, in relation to computer science, and the foundational problems of machine ethics are required. Based on these the students are enabled to identify and denominate ethical conflicts and dilemmas by themselves and derive standpoints, based on research of related argumentations and their critical analysis.

The interdisciplinary nature of the course topic is reflected by an iterative, alternating succession of lectures and group discussions (cf. Figure 2). In addition, so-called flipped or inverted classroom [13] sessions are integrated as well. After an initial introduction of the course concept and outline, we start with a discussion of selected case studies.

After an initial Introduction of the course concept and outline, we start with a discussion of selected case studies. Thankfully, a working group of the *Gesellschaft für Informatik*, i.e., an expert society for Computer Science in German-speaking Countries, regularly makes examples of ethical dilemmas freely available [14,15]. Based on a preselection of case studies, groups of students are formed and engage in a group discussion.

Finally, the results are shared in plenum. Initially, being plunged in at the deep end gives students an intuitive impression of the kind of ethical reasoning that is required within the course. Afterwards, foundational ethical principles and theories are introduced. Here, we focus a.o. on *virtue ethics*, *contractualism*, *sentimentalism, utilitarianism*, *deontological ethics* and *contract theory* according to Rawls [1,16]. the subsequent group works, student discussion explores the applicability of these theories in the context of computer science applications. Subsequently, the concept machine ethics and the challenge of anticipatory technology assessment are introduced. The later concept is supplements by introducing agent-based modeling and system dynamics modeling approaches [17].

This perspective allows students to reason about the causal inter-relations between technological components and stakeholders. The machine ethic perspective is consolidated by the identification of ethical agents and corresponding challenges in group works. Later, the technical foundations are given. Machine ethics requires a basic understanding of development approaches to autonomous system entities. The autonomous operation of machines requires localized reasoning capabilities and the ability to adapt at run-time. Thus, software agents, agent architectures, and approaches for developing adaptive systems are introduced [18,19]. We particularly address the development of agents following the Belief-Desire-Intention architecture [18]. This architecture allows to highlight challenges in embedding reasoning processes of individual technical components.
The elective module is available to undergraduates at various points in their academic programme. Therefore, the creation of the concluding written tasks commences with an exploration of scientific writing and literary investigation. Two inverted class sessions maintain this focus by encouraging close analysis of specific literature. These sessions contemplate (1) measures for compelling machine regulations [20] and (2) incorporation of ethical deliberations in commercial growth frameworks, specifically those relating to Big Tech [21]. Students quickly identify limitations and possibilities in using artificial intelligence generated texts.

Students select, after consultation and advice from the lecturers, their final assignment topic for the final grading. The topic selection clearly reflects current major topics of the ethic discourse in computer science research. Here, we structure the identified subject areas. The fundamental aspects of machine ethics are a recurring theme. These range from the applicability of ethical theories, e.g., Shintoism and Confucianism to granting rights to technical entities or impacts on Human Machine interaction. Another larger cluster concerns the manifestation of reasoning capabilities in physical entities, i.e., robots or vehicles. Differing Application areas are considered, prevalent are discussions of governmental-run applications and Healthcare systems.

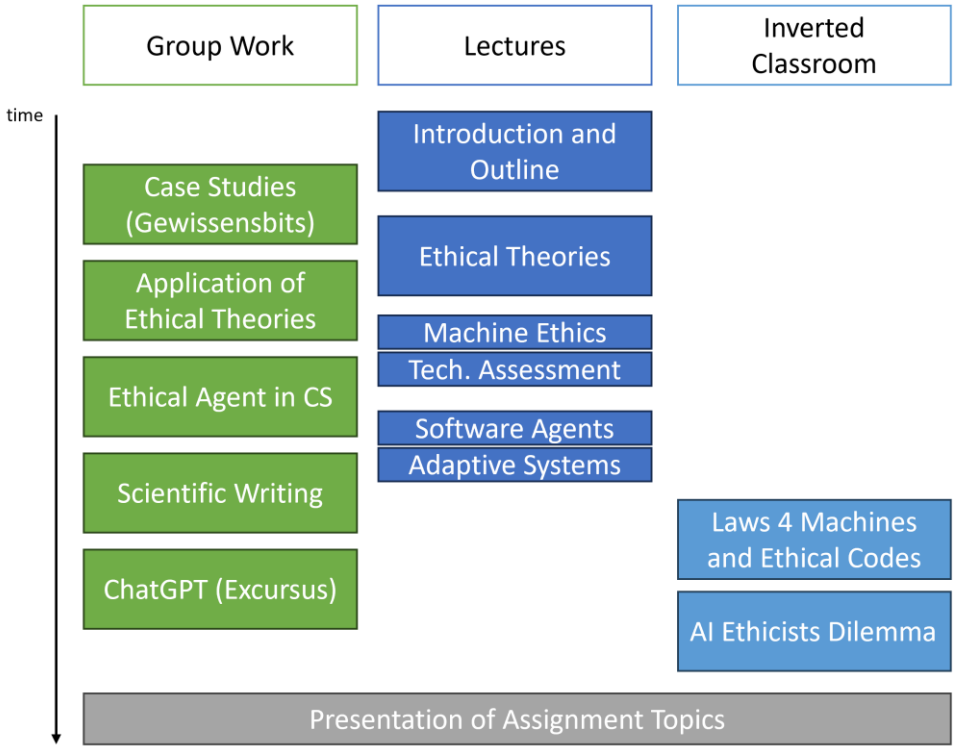

Figure 2. Course outline, lectures, group discussions and inverted classrooms alternate are used to facilitate interdisciplinary reflection and exchange among students.

Reflection of the students' choice of topics

The overview below (Figure 3) shows a selection of the students' homework topics from the two seminar courses. The students mainly chose the topics themselves or narrowed them down with our support. As previously mentioned, numerous fundamental topics and questions were presented, with many of them concentrating on AI-related subjects like "Should robots lie?" or "To what extent should we investigate AI?" Furthermore, the students were also interested in topics that dealt with the state's use of intelligent systems, automatic decision making or intelligent decision support, such as the use of automatic weapons systems, the introduction of AI into the work of courts or also in law enforcement. Questions on the differentiated use of autonomous systems in the care and health sector were also of interest to the students, such as "Use of AI in psychology/psychotherapy & ethical implications", "The use of AI in hospitals and the associated consideration of liability" or "Applicability of basic medical ethical principles to nursing robots". Some topics also dealt with the question of the actor status of intelligent systems and the increasing interrelationship at" eye level" between humans and their machines, e.g. "Azuma Hikari: A relationship with a hologram without restrictions?", "Sex robots in our society", Dax robots - The supplier robot of the future?" and "Intelligent astronaut assistants: The future of the space mission". Finally, there were some term papers that dealt with the differently positioned danger and risk potential of intelligent systems like "Tiktok - The algorithm that gets us hooked" or "How racist can machines be?". All term paper topics were presented by the students in five-minute short presentations within the seminar group, briefly discussed as well as feedback from the lecturers and fellow students on the topic and term paper structure. The students then had six weeks to complete their term paper.

Topic cluster 1 Fundamental issues	Topic cluster 2 State-deployed AI	Topic cluster 3 AI in care & health	Topic cluster 4 AI & robots as actors	Topic cluster 5 Dangers of automated decision making
- Social Companions - Should Robots Lie? - Robot rights - How far should we explore AI? - Confucian values in robot ethics - To what extent is Shintoism helpful/useful in machine ethics? - Human ethics for non-human actors? - Consent of humans in interaction with assistance systems - Social classification of AI-generated art	- To what extent should autonomous war robots be used in military conflicts? - Automated Weapon Systems, FCAS - Police robots - Artificial intelligence in law enforcement - Automation of the courts - risks for the justice system?	- Use of AI in Psychology/Psychotherapy & Ethical Implications - The use of AI in hospitals and the associated consideration of liability. - Applicability of basic medical ethical principles to care robots	- Azuma Hikari: A relationship with a hologram without restrictions? - Sex robots in our society - Daxrobots - The supplier robot of the future? - Intelligent astronaut assistants: The future of the space mission	- Tiktok - The algorithm that gets us hooked - How racist can machines be?

Figure 3. Selected topics of term papers

4 CONCLUSIONS

What have we learned during the first two classes of this elective course? What has been successful so far and what can be improved in the future? Have the students achieved their learning objectives and how frequent is the interdisciplinary dialogue between the teachers' respective perspectives? Is there a tangible added value as a result? If so, how have we established it? The following hints provide an outlook on this and the further development of the seminar.

Overall, the feedback from both seminar groups was encouraging. Specifically, the students appreciated the seminar topic and the flexible format, which allowed for lively discussions and individual topic selection. Nevertheless, some students expressed skepticism about the ethical frameworks employed, raising concerns about practical application in their future professional lives. We were unable to adequately address the transfer of knowledge between ethical considerations, reflections on guidelines and concrete technical applications with numerous development constraints during the course as it exceeded the seminar's scope. Nevertheless, we commend the students for their high level of commitment, and we were impressed by their reflective and ambitious term paper topics.

A prospect for future work is complementing the theoretical reasoning, presented here, with a more detailed analysis and understanding of implementation aspects (e.g., discussed in [23]). One approach to pursue this is engaging students in project works, where groups build examples of systems which have ethical impact. These impacts are evident when building systems in close vicinity to humans, e.g., companions or smart home applications. We want to explore utilising infrastructures at HAW Hamburg, e. g. the LIVING PLACE lab [4] where an overlay of software agents could control appliances and interact with human inhabitants. In addition, it is particularly attractive to strive for embedding ethical educational modules throughout the Computer Science curriculum, as proposed in [7].

ACKNOWLEDGEMENTS

We would like to thank all the students who took part in the first iterations of the courses described, for their active participation, open discussions and insightful feedback.

REFERENCES

[1] C. Misselhorn, Grundfragen der Maschinenethik. Reclams Universal-Bibliothek, Reclam Verlag, 2018.

[2] M. Anderson and S. L. Anderson, Machine ethics: Creating an ethical intelligent agent. AI magazine, 28(4), pp. 15-26, 2007.

[3] K. Miller, Integrating Computer Ethics into the Computer Science Curriculum. Computer Science Education, 1:1, pp. 37-52, 1988.

[4] C. Fiesler, N. Garrett and N. Beard. What do we teach when we teach tech ethics? A syllabi analysis. In Proceedings of the 51st ACM technical symposium on computer science education, pp. 289-295, 2020.

[5] J. S. Saltz, N. I. Dewar and R. Heckman. Key concepts for a data science ethics curriculum. In Proceedings of the 49th ACM technical symposium on computer science education, pp. 952-957, 2018.

[6] O. Bendel, Wozu brauchen wir die Maschinenethik? Handbuch Maschinenethik, Springer, pp. 13-32, 2019.

[7] E. Burton, J. Goldsmith and N. Mattei, How to teach computer ethics through science fiction. Commun. ACM 61, pp. 54–64, 2018.

[8] C. Canca, Operationalizing AI ethics principles. Commun. ACM 63, pp.18–21, 2020,

[9] B. J. Grosz, D. G. Grant, K. Vredenburgh, J. Behrends, L. Hu, A. Simmons, and J. Waldo, Embedded EthiCS: integrating ethics across CS education. Commun. ACM 62, 8, pp. 54–61, 2019.

[10] D. Horton, S. A. McIlraith, N. Wang, M. Majedi, E. McClure and B. Wald, Embedding Ethics in Computer Science Courses: Does it Work. In Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2022), Providence, RI, USA. ACM, New York, NY, USA, 7 pages. 2022.

[11] T. S. Goetze, Integrating Ethics into Computer Science Education: Multi-, Inter-, and Transdisciplinary Approaches. In Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2023). Association for Computing Machinery, New York, NY, USA, pp. 645–651. 2023.

[12] I. Mezgár, J. Váncza, From ethics to standards – A path via responsible AI to cyber-physical production systems. Annual Reviews in Control 53, pp. 391–404, 2022,

[13] M. L. Maher, C. Latulipe, H. Lipford and A. Rorrer, Flipped Classroom Strategies for CS Education. Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15). Association for Computing Machinery, pp. 218–223, 2015.

[14] C. B. Class, W. Coy, T. Dührsen, B. Kees, C. R. Kühne, C. Kurz, O. Obert, R. Rehak, G. Schiedermeier, C. Trinitis, S. Ullrich and D. Weber-Wulff, Gewissensbits Fallbeispiele zu Informatik und Ethik, Accessed 22 July 2023. Retrieved from https://gewissensbits.gi.de/

[15] C. B. Class, W. Coy, C. Kurz, O. Obert, R. Rehak, C. Trinitis, S. Ullrich and D. Weber-Wulff (Eds.), Gewissensbisse - Fallbeispiele zu ethischen Problemen der Informatik. transcript Verlag, 2023.

[16] M. H. Werner, Einführung in die Ethik. Metzler Verlag. Berlin, 2021.

[17] P. Barbrook-Johnson, A. S. Penn, Systems Mapping How to build and use causal models of systems, Palgrave Macmillan, Open Access, 2022.

[18] M. Wooldridge, An Introduction to MultiAgent Systems, second edition, Wiley, 2009

[19] M. Wooldridge, The Road to Conscious Machines, A Pelican book, 2020

[20] M. Funk, Welchen Regeln und Gesetzen müssen Maschinen folgen? (Bedeutung 4). In: Roboter- und KI-Ethik. Springer Vieweg, Wiesbaden, 2022.

[21] H. S. Sætra, M. Coeckelbergh, J. Danaher, The AI ethicist's dilemma: fighting Big Tech by supporting Big Tech. AI Ethics 2, pp. 15–27, 2022.

[22] OpenIA, Introducing ChatGPT, Accessed 23 July 2023. Retrieved from openai.com

[23] D. Trentesaux, S. Karnouskos, Engineering ethical behaviors in autonomous industrial cyber-physical human systems. Cogn Tech Work., 2021.

[24] Hamburg University of Applied Sciences, Labor “Living Place”, Accessed 10. Sept. 2023. Retrieved from https://livingplace.haw-hamburg.de/