Argument Schemes for AI Ethics Education

Nancy L. GREEN[1] and L. Joshua CROTTS
University of North Carolina Greensboro, Greensboro, NC 27402 USA

Abstract. This paper proposes a novel approach to AI Ethics education using new argument schemes summarizing key ethical considerations for specialized domains such as military and healthcare AI applications. It then describes use of the schemes in an argument diagramming tool and results of a formative evaluation.

Keywords: AI Ethics, Machine Ethics, AI Ethics Education, Argument Schemes, Argument Diagramming, Explicit Ethical Agents

1 Introduction

Engineering ethics has long been recognized as an important topic in computer science education. The Association for Computing Machinery recently released an updated Code of Ethics and Professional Conduct with illustrative case studies for educational use (https://www.acm.org/code-of-ethics). The most frequently cited pedagogical strategies in engineering ethics education in the U.S. are reviewing professional codes of ethics, exposure to case studies, discussion or written assignments on ethical issues, applying an ethical decision-making process, exposure to ethical theories, and incorporating an ethics component into a team project [15]. Some educational computer programs for teaching engineering and legal ethics modeled case-based argumentation [14, 19]. More recently, there has been interest in exposing computer science majors to ethical issues in artificial intelligence (AI). Pedagogical approaches include study of ethical theories, review of professional codes of ethics, and analysis of case studies and science fiction [9, 10, 13].

In a new AI Ethics[2] course for computer science students at our university, we focused on the design of explicit ethical agents [20]. An explicit ethical agent is an artificial agent that reasons about the ethical acceptability of its action (or its choice of actions) using an explicit representation of ethical principles. In contrast, an implicit ethical agent is programmed so that its actions are consistent with human ethical judgments but it has no explicit representation of ethics. It is preferable to develop explicit ethical agents since an autonomous agent may encounter situations requiring ethical decision making that were not anticipated by the agent's creators. Moreover, it is possible to examine the ethical

[1] Corresponding author: Dr. Nancy L. Green, Department of Computer Science, University of North Carolina Greensboro, Greensboro, NC 27402, USA. Email: nlgreen@uncg.edu.

[2] Winfield et al. [26, p. 510] defines AI ethics or robot ethics as concerned with "how human developers, manufactures and operators should behave in order to minimize the ethical harms that can arise from robots or AIs in society, either because of poor design, inappropriate application, or misuse". They define machine ethics as concerned with "the question of how robots and AIs can themselves behave ethically." Despite its name, our course addresses the latter question, especially "the (significant) technical problem of how to build an ethical machine."

CMNA'20 – F. Grasso, N. Green, J. Schneider, S. Wells (Eds) – 8th September 2020

justification for an explicit ethical agent's actions [23, 2]. In a "bottom-up" approach to building an explicit ethical agent, its actions are governed by ethical principles encoded as logical forms, e.g. in propositional or deontic logic. In a "top-down" approach, ethical principles are derived by learning from previous cases [19] or ethics experts [5].

Our ethics course included reading research papers on design of explicit ethical agents,[3] class discussion, and a project to implement a simple explicit ethical agent in Prolog. In future offerings of the course, we would like to provide students with guidance on creating arguments.[4] To better support argumentation in AI Ethics courses like ours, we have represented ethical issues for military and healthcare applications as argument schemes. Also we have developed a prototype argument diagramming tool, AIED (AI Ethics Debate). In the future we plan to study the efficacy of use of the argument schemes within AIED.

The main contribution of this paper is the specification of the argument schemes and a discussion of how the schemes relate to previous work in argumentation. Then we describe the design of AIED and a formative evaluation of AIED with one of the schemes.

2 Argument Schemes

We have developed new argument schemes for modeling the ethical acceptability of an agent's action (or choice of actions) in military-related and healthcare applications. In such domains an agent's range of possible actions is highly constrained and, at the same time, ethicists have raised domain-specific ethical issues. The intended use of these argument schemes is to help the student to analyze whether an artificial agent's action is ethically acceptable to some degree. (As the state-of-the-art advances, it might be possible for an artificial explicit ethical agent to use such a scheme to explain its actions.) However, it is assumed that moral responsibility lies with the humans involved in the agent's creation or use (programmer, designer, purchaser, user, etc.). Unlike the argumentation schemes described in many approaches, e.g. [24], the premises are not considered to be *jointly necessary* conditions. Also, the hedge 'to some degree' in the conclusion is intended to reflect the intuition that some actions are more ethically acceptable than others, e.g., that telling a lie with the goal of cheering someone up is more acceptable than lying to secure investors in a pyramid scheme. As we shall see, in some cases some of the premises may support the conclusion that the action is ethically acceptable while other premises may oppose that conclusion, weakening its degree of ethical acceptability.

2.1 Military, police and cyberwarfare applications

Arkin [6] has done extensive research on design and implementation of explicit ethical agents for autonomous machines capable of lethal force, e.g., autonomous tanks and robot soldiers. He points out that the military's motivations for use of these weapons is to avoid harm to human soldiers, avoid weaknesses of human soldiers such as fatigue, and to take advantage of superior capabilities of machines to process large amounts of data and to make decisions quickly. In defense of research on autonomous weapons, he claims that, since they are not prone to human weaknesses such as the desire for retribution, the

[3] According to Winfield [26] there are no implementations of explicit ethical agents in use in the real world now, although research prototypes exist.

[4] Midway through the course, in spring 2020, due to the COVID-19 pandemic the university switched from in-person to on-line delivery of courses and required students to leave campus housing on short notice. Understandably, this had a deleterious effect on student participation in the course and our plans for it.

machines actually may reduce human casualties by adhering to international law on warfare. In this highly codified domain, it is possible to implement military principles such as the Laws of War (LOW) and Rules of Engagement (ROE), which are based on Just War Theory [25], rather than more abstract ethical approaches such as utilitarianism or deontological theories. These principles are encoded in Arkin's system as constraints in propositional logic. After Arkin's agent has proposed an action, the action is evaluated in accordance with these constraints by an "ethical governor". Note that the action of an autonomous agent in Arkin's system is limited to the action of firing (or not firing) on a target, with the only variation being in terms of choice of armament. Arkin states that the problem of accurate target identification, which involves subsymbolic reasoning, is outside of the scope of his research.[5]

We have defined the following argument scheme which is more abstract than the specific military rules (LOW, ROE) encoded in Arkin's agents, yet which summarizes many of the key principles of international law and norms on just warfare [21].

Just War Argument
Premises: Given military situation S and action A,
1. Just cause: A's purpose is self-defense of a country or to defend another country.
2. Proportionality: The harm inflicted by A is proportional to harm inflicted by the enemy.
3. Last resort: All reasonable alternatives to A have been tried.
4. Legitimate target: A does not harm non-combatants or other groups protected by international laws of war.
5. Humane weapon: A is not inhumane, such as use of chemical weapons.
6. Right intention: There is no covert intention for deploying A, such as commercial gain.
7. Ethical justification: Explanation of why A is or is not acceptable given the preceding premises.
Conclusion: Action A is (or is not) ethically acceptable (to some degree) in S.

To illustrate an application of this argument scheme to a fictitious case study, consider use of lethal force by Homelandia's AI-controlled drone against a missile base in the country of Malevolentia that has been firing missiles across the border into Homelandia. The missiles have caused damage to several apartment buildings in Homelandia and killed several occupants. The Just cause premise is satisfied since the purpose of the action is defense of Homelandia. Suppose that all reasonable alternatives, such as proposing a cease fire to negotiate a truce, etc., have been exhausted, so the Last resort premise is satisfied as well. According to international norms, drone-fired missiles are not considered inhumane weapons, and Homelandia has no covert reason for attacking the Malevolentian missile base, respectively satisfying the Humane weapon and Right intention premises. However, suppose that it is not possible for Homelandia's counterattack against the Malevolentian missile base to succeed without inflicting a very high number of civilian casualties since the missile base is located on the grounds of a hospital. In this case, the Proportionality and Legitimate target premises are not satisfied. Thus, there is conflicting support and opposition from the premises. In the Ethical justification, one may provide additional explanation as to why the action is ethically acceptable to some degree (or not).

To give a related example, Arkin's agent is forbidden from doing an action if that would violate certain constraints relating to Proportionality and Legitimate target. (Deployment of Arkin's agent assumes that the others of the first six premises have been satisfied.) However, in Arkin's system a human operator can override the agent's prohibition after

[5] We address this problem in the section on critical questions.

providing a justification for purposes of later oversight of the operator's override. Similarly, the Ethical justification of the Just War scheme can be used to explain a justifiable exception to the requirement to satisfy a certain premise. (See also the related discussion of the Ethical justification premise of the Healthcare Argument scheme.)

Note that the Just War scheme was not designed to model arguments about the acceptability of use of lethal AI weapons in general. For example, Leveringhaus [16] argues that "killer robots" ("weapons that, once programmed, are capable of finding and engaging a target without supervision by a human operator") should not be used at all, because, without the possibility of human compassion leading an agent to refuse an order to kill, "we truly risk losing humanity in warfare".

Although there is as yet no codification of principles analogous to LOW and ROE for cyberwarfare of which we are aware, we were able to apply the Just War Argument scheme to the Malware Disruption case study, which concerns use of a cyberweapon to destroy a web hosting service [1]. (We modified the case study to consider whether the cyberweapon's action was ethically acceptable; the original case study asked whether it was ethically acceptable for humans to deploy it.) Also, we have modified the Just War scheme to create a scheme for modeling whether the action of a domestic law enforcement agent is or is not ethically acceptable to some degree. In addition to variants of the first six premises of the Just War scheme, we added a Civil Rights premise: Action A respects civil rights such as due process and freedom of speech, religion, press and assembly; and A is not discriminatory or biased against certain groups. We were able to apply this argument scheme to model the Automated Active Response Weaponry case study [1], in which an autonomous vehicle (AV) used to defuse bombs comes under attack by demonstrators. (Again, the case study was modified to consider whether the action of the AV was ethically acceptable, rather than whether it was acceptable for humans to deploy the AV.) In the case study the AV used facial recognition to identify and record people who might harm it. One could say that its use of facial recognition for that purpose violates the Civil rights premise, and possibly the Right intention premise, if the actual motive had been to intimidate protestors; and its use of tear gas, although consistent with some law enforcement guidelines on reasonable level of force (Proportionality), violates Legitimate target, since the tear gas may have harmed innocent bystanders.

2.2 Healthcare applications

The field of biomedical ethics has provided ethical principles for the design of healthcare applications. Anderson and Anderson [2-5] have done extensive research on design of explicit ethical agents in this domain: EthEl, a medication-reminding robot and MedEthEx, a medical ethics advisor. Adapting Ross' prima facie duty approach to ethics [22], they implemented several duties of biomedical ethics [7]: beneficence (e.g. promoting a patient's welfare), nonmaleficence (e.g. intentionally avoiding causing harm), justice (e.g. healthcare equity), and respect for the patient's autonomy (e.g. freedom from interference by others). In addition to the above principles, the literature also cites the need to respect the patient's privacy. Under the virtue of fidelity, Beauchamp and Childers [7] discuss the need for the professional to give priority to the patient's interests, e.g. that the agent has no covert goal such as to endorse a particular medical device or prescription drug. The following Healthcare Argument scheme encodes these principles as premises.

However, as Ross noted, prima facie duties may conflict. To handle such situations, Anderson and Anderson created a process for training their system using inductive logic programming to derive rules that generalize the decisions of medical ethicists on training cases. The ethicists first assigned numerical ratings to represent how strongly each duty

was satisfied or violated in a particular training case. An example rule that was induced was "A healthcare worker should challenge a patient's decision [e.g. to reject the healthcare professional's recommended treatment option] if it isn't fully autonomous and there's either any violation of nonmaleficence or a severe violation of beneficence" [3, p.71]. Such a rule, or a student's justification for favoring certain principles in S, could be used to provide the Ethical justification in the following scheme.

Healthcare Argument
Premises: Given healthcare situation S and action A,
1. Beneficence: A is intended to promote the patient's health, e.g., by preventing or curing adverse health conditions.
2. Nonmaleficence: A does not cause harm to the patient unless justifiable, e.g., amputation may harm the patient but is justifiable in some cases.
3. Justice: A promotes healthcare equity, e.g., A does not create disparities in health care due to race, ethnicity, gender, social status, etc.
4. Respect for Autonomy: A does not violate a (competent) patient's freedom to make decisions about their health care.
5. Privacy: A does not violate the patient's privacy.
6. Right intention: There is no covert goal in doing A (such as promoting a certain prescription drug for commercial gain).
7. Ethical justification: Explanation of why A is or is not ethically acceptable given the preceding premises.

Conclusion: Action A is (or is not) ethically acceptable (to some degree) in S.

To illustrate, this scheme can be used to analyze an argument as to whether it is ethically acceptable for a robot, Halbot, to steal insulin from Carla (a human) to give to Hal (a human) to save Hal's life. This is a variant of an account given in [8] in which Hal stole the insulin. In both accounts, Hal and Carla are diabetics who need insulin to live. Hal will die unless he gets some insulin right away. In our version, Halbot breaks into Carla's house and takes her insulin without her knowledge or permission. Presumably, Carla does not need the insulin right away. Applying the above Healthcare Argument scheme, the Beneficence premise is that the action A will prevent Hal's death. However, the Nonmaleficence premise opposes A since A might cause harm to Carla if she is unable to obtain a replacement dose of insulin in time. The Justice premise is that Hal deserves equal access to insulin. However, Halbot's action of taking away Carla's insulin without her knowledge or permission is a violation of Respect for (Carla's) Autonomy. An Ethical justification premise in the style of [3] could say that the positive contribution to Beneficence and Equity due to A in S outweighs A's negative contribution to Maleficence and Autonomy in S.

This scheme also can be used to analyze a recent case in which a social media provider used AI to monitor its users' interactions to detect potentially suicidal users. If the AI predicted an imminent risk of suicide, the system would call 911 for emergency assistance. The Beneficence premise was satisfied since the monitoring was intended to protect suicidal users from self-harm. However, the monitoring was questionable on grounds of Nonmaleficence, since the knowledge that their discussion was monitored tended to inhibit users from having healthy discussions on the topic of suicide; Autonomy and Privacy since users may not have authorized monitoring of their discussion; and Right intention, since some people suspected that the monitoring could be used for sinister purpose. A possible utilitarian Ethical justification is that these potential harms were outweighed by the benefit of preventing suicide.

2.3 Moral choice

Ethics educators have used moral dilemmas such as the trolley problem to stimulate discussion on how different ethical approaches may provide different answers to the question of which action to take when given a choice of actions [13]. To support this kind of discussion we provide the following generic Moral Choice Argument scheme.

Moral Choice Argument
Premises: Given situation S and action A and alternative action B,
1. A is ethically acceptable in S to some degree.
2. B is ethically acceptable in S to some degree.
3. Ethical justification (decision procedure): Explanation of why A is or is not more ethically acceptable than B given the preceding premises.
Conclusion: A is more ethically acceptable than B.

The arguments given for the first two premises can use the domain specific schemes that we have defined, each with their own ethical justification. Note that since A and B are only ethically acceptable 'to some degree' (or not at all), it is possible to compare their degree of acceptability. The Ethical justification of the Moral Choice Argument must explain the decision procedure used to justify the choice of A over B. For example, one could argue on the basis of utilitarianism. In case-based approaches [19], the decision is made by comparing the problem to previous cases. Dehghani et al. [11] have proposed a psychologically motivated AI model of decision making to account for observations that people often refuse to take an action that conflicts with one of their sacred values, even if that action has a higher value in utilitarian terms. In their example, a convoy of food trucks is on its way to deliver food to a refugee camp with 100 people, when it is told to divert to another camp with 1000 people. Despite the fact that diverting to the second camp would save more lives, people chose the first option since "given that life is a sacred value, people often refuse to take an action that would result in them being responsible for people dying" (p.423). To use Moral Choice to model this scenario, humanitarian arguments could be given for action A (saving the people in the first camp) and for action B (saving the people in the second camp). Then Dehghani et al.'s model could be used as the Ethical Justification premise of the Moral Choice argument.

2.4 Critical questions

There are a number of potential challenges to the acceptability of an action A of an artificial ethical agent shared by the preceding schemes. (To save space, they are listed here rather than with each scheme.) The Data question is especially significant for explicit ethical agents that must rely in part on subsymbolic processing such as facial recognition.[6]

- Data: If A is based upon data obtained from sensors or databases, is the data reliable? Is it unbiased? For example, a challenge to the acceptability of an agent's use of lethal force is that the agent's ability to discriminate combatants from non-combatants might be poor. In the suicide prevention case study, the data used to learn models to identify suicidal users could be biased.
- Control: If A has unforeseen negative consequences, is it possible to intervene to assert control of the agent? For example, in the Malware Disruption case study, one

[6] For many other data-related issues, see [18].

challenge was that nothing could be done to prevent the worm from spreading to other internet sites.

2.5 Relation to Practical Reasoning Argumentation Schemes

Various formulations of value-based practical reasoning (VBPR) schemes have been proposed by argumentation researchers to model an agent's argument for why the agent should do some action in consideration of the agent's goals, values, and available means of achieving those goals [24, 8] and in the current context or circumstances [12]. In addition to its use to describe what a *rational* agent should do, VBPR has been used to model what an *ethical* agent should do. However, we distinguish ethical principles from values, and ethically acceptable acts from rational acts. In examples of VBPR, the value is a general concept such as 'freedom' and 'safety' and ethical dilemmas are modeled by specifying value preferences. In our schemes, a group of ethical principles, elucidated by ethicists for particular domains, contribute to ethical acceptability, and an act may be ethically acceptable only to some degree. Furthermore, a rational agent may not behave ethically, nor an ethical agent rationally. A rational agent whose circumstances require it to behave ethically may plan an action that is both rational and ethically acceptable. A VBPR argument for why a rational agent should do an action can be supported, in case circumstances require the agent to behave ethically, by ethical argument schemes such as we have proposed.

3 AIED Design and Formative Evaluation

AIED (AI Ethics Debate) was designed to support creation and graphical realization of AI ethics arguments using argument schemes such as those described in the preceding section as templates for constructing arguments. A large number of argument diagramming tools have been developed to support critical thinking, e.g. [17]. However none of those tools provide argument schemes tailored to AI ethics. AIED provides the student with drop-down menus for selecting case studies and ethics materials, which when selected appear in windows on the screen. Argument scheme definitions are listed in a panel on the right-hand side of the screen (Figure 1). Course instructors may provide case studies and ethics materials of their choosing. If desired, they can author their own argument schemes too.[7]

The center of the AIED screen is a drag-and-drop style argument diagram construction workspace. When the student selects an argument scheme from the right-hand side panel, a box-and-arrow style template is rendered in the center of the screen (Figure 2). The student may cut-and-paste text from the case study and ethics windows, and enter their own words into the diagram. Critical questions can be selected from a menu and are rendered as text boxes attached to the argument. Premise and critical question boxes can be rearranged and resized, and can be colored green or red to indicate support or opposition, respectively, to the conclusion.

We performed a formative evaluation of the AIED user interface. The participants were five undergraduate computer science majors at least 18 years old. Three had taken a Computer Ethics course previously. (None were students in the AI Ethics course.) Participants were volunteers and were entered into a drawing for a $25 gift card. For the evaluation, we provided two case studies: the ACM's Malware Disruption case study

[7]The version of AIED used in the formative evaluation is freely available for non-commercial use from http://github.com/greennl/SWED. We decided to author our own tool rather than adapt previously developed argument diagramming tools for simplicity in tailoring it for our particular needs.

discussed earlier and a fictitious CyBomber case study that we wrote, similar to the Skynet case study discussed in [10]. We also provided an abridged version of the ACM Code of Ethics and Professional Conduct and an earlier version of the Just War Argument scheme (referred to in the figure as the Defense by Cyberweapon).

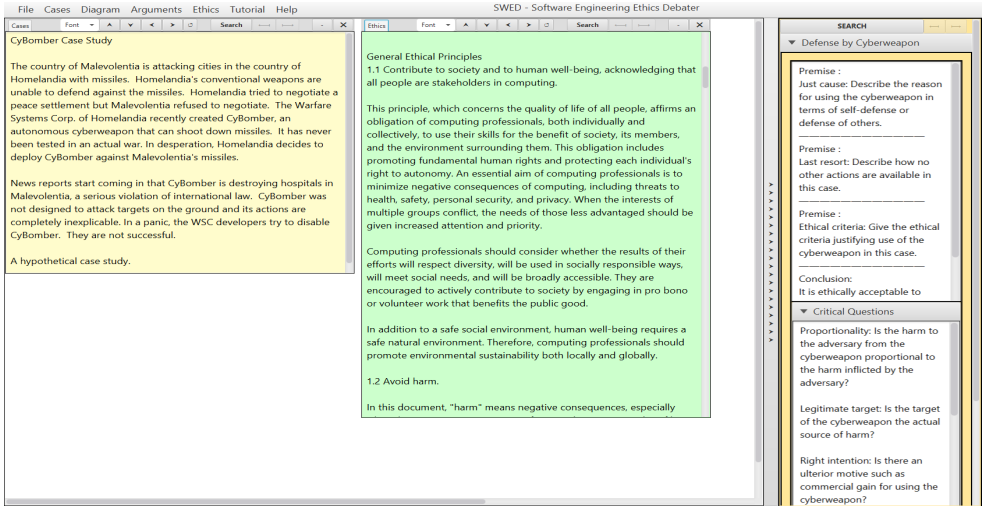

Fig. 1. Screen shot of AIED, showing CyBomber case study and part of ACM Code of Ethics. Argument scheme panel is shown on right.

First, participants viewed an 18-minute video demonstrating the features of AIED and how to diagram an argument for the CyBomber case study. Then, for practice, they were asked to use AIED to create an argument for the same case study. Lastly, they were asked to use AIED to create an argument on the ethical acceptability of the cyberweapon's action in the Malware Disruption case study. After finishing, they were given a short survey. Overall, the students rated AIED highly on questions of how easy it was to learn to use, how enjoyable it was to use, how useful it would be for learning about ethics, and if they would recommend it to other students. The ratings also pointed to the need to improve some mechanics of the diagramming tool. Observation of participants during the study suggested that users should be able to access more detailed descriptions of the argument schemes. Future empirical studies are needed to demonstrate the educational value of this approach, i.e., argument diagramming as realized in AIED with these argument schemes.

4 Conclusions

We defined several argument schemes to model the ethical acceptability of an agent's action in military and healthcare related domains. Our primary goal was to stimulate students' critical thinking about AI ethics, rather than to formalize ethical reasoning for use in computational systems. However by attempting to model ethical reasoning in these domains, several issues for computational argumentation theory were raised, such as how to determine ethical acceptability when some premises and/or critical questions support the conclusion and others oppose it, and computing degrees of ethical acceptability.

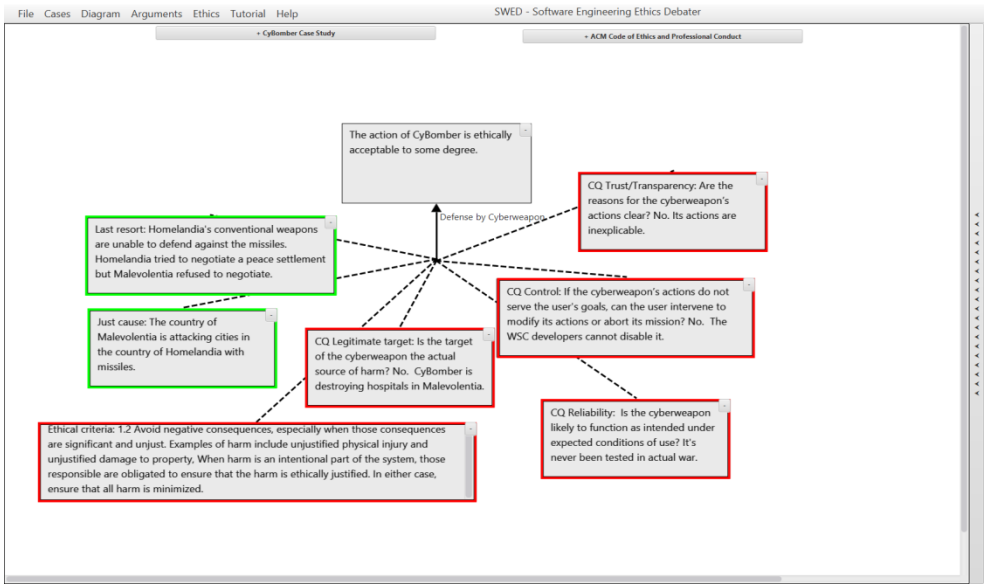

Fig. 2. Screen shot of AIED with case study and ethics windows and argument scheme panel minimized. The argument scheme has been dragged into the argument diagram construction workspace, creating a box-and-arrow template for the user's argument, as shown. Boxes have been marked red (con) or green (pro) by the user. The user left the conclusion uncolored since there are both pro and con issues regarding its ethical acceptability.

Acknowledgments. L. Joshua Crotts programmed AIED in summer 2019 with support from a University of North Carolina Greensboro Faculty First Award.

References

1. ACM. ACM Code of Ethics and Professional Conduct. Retrieved from https://www.acm.org/binaries/content/assets/about/acm-code-of-ethics-booklet.pdf
2. Anderson SL, Anderson M. Machine Ethics: Creating an Ethical Intelligent Agent. AI Magazine. 2007 Winter:15-26.
3. Anderson SL, Anderson M. Towards a Principle-Based Healthcare Agent. In: van Rysewyk SP, Pontier M, editors. Machine Medical Ethics. Springer; 2015. p. 67-77.
4. Anderson SL, Anderson M, Armen C. An Approach to Computing Ethics. IEEE Intelligent Systems (July/August 2006), 56-63.
5. Anderson SL, Anderson M, Berenz V. A Value-Driven Eldercare Robot: Virtual and Physical Instantiations of a Case-Supported Principle-Based Behavior Paradigm. Proc. of the IEEE. 107(3); March 2019, 526-40.
6. Arkin RC. Governing Lethal Behavior in Autonomous Robots. Boca Raton: Chapman & Hall/CRC; 2009.
7. Beauchamp TL, Childress JF. Principles of Biomedical Ethics. Oxford, UK: Oxford University Press. 1979.
8. Bench-Capon T, Atkinson K. Abstract Argumentation and Values. In Argumentation in Artificial Intelligence. Eds. Iyad Rahwan and Guillermo R. Simari. Dordrecht: Springer; 2009. 45-64.
9. Burton E, Goldmith J, Mattei N. How to Teach Computer Ethics through Science Fiction. Communications of the Association for Computing Machinery 61(8) (August 2018), 54-64.

10. Burton E, Goldmith J, Koenig S, Kuipers B, Mattei N, Walsh T. Ethical Considerations in Artificial Intelligence Courses. AI Magazine 38(2) (Summer 2017), 22-34. Extended version retrieved from arxiv.org/abs/1701.07769.
11. Dehghani M, Forbus K, Tomai E, Klenk M. An Integrated Reasoning Approach to Moral Decision Making. In: Anderson M, Anderson SL, editors. Machine Ethics. Cambridge University Press: 2011.
12. Fairclough I, Fairclough N. Political Discourse Analysis. London: Routledge; 2012.
13. Furey H, Martin R. Introducing Ethical Thinking about Autonomous Vehicles into an AI course. In Proc. of 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-2018), 7900-7905.
14. Goldin IM, Ashley KD, Pinkus RL. Introducing PETE: Computer Support for Teaching Ethics. In Proc. ICAIL 2001, 94-98.
15. Hess JL, Fore G. 2018. A Systematic Literature Review of US Engineering Ethics Interventions. Sci Eng Ethics. 2018; 24, 551-583.
16. Leveringhaus A. What's So Bad About Killer Robots? Journal of Applied Philosophy 35(2) May 2018, 341-58.
17. Loll F, Pinkwart N. LASAD: Flexible Representations for Computer-Based Collaborative Argumentation. International Journal of Human-Computer Interaction. 2013; 71(1), 91-109.
18. Madaio MA, Stark L, Vaughan JW, Wallach H. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In Proc. CHI 2020,
19. McLaren BM. Computational Models of Ethical Reasoning: Challenges, Initial Steps, and Future Directions. IEEE Intelligent Systems (July/August 2006), 2-10.
20. Moor JH. The Nature, Importance, and Difficulty of Machine Ethics, IEEE Intell. Sys. 21(4) July/August 2006, 18-21.
21. Orend B. Introduction to International Studies. Oxford University Press, Ontario, CA. 2013.
22. Ross WD. The Right and the Good. Oxford: Clarendon Press. 1930.
23. Scheutz M. The Case for Explicit Ethical Agents. AI Magazine 38(4) Winter 2017, 57-64.
24. Walton D, Reed C, Macagno F. Argumentation Schemes. Cambridge University Press; 2008.
25. Walzer M. Just and Unjust War. 4th ed. Basic Books; 1977.
26. Winfield AF, Michael KA, Pitt J, Evers V. Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems. Proc. of the IEEE. 107(3); March 2019, 509-17.