Macro Ethics for Governing Equitable Sociotechnical Systems

Blue Sky Ideas Track

Jessica Woodgate
The University of Bristol
Bristol, UK
jesswoodgate@yahoo.co.uk

Nirav Ajmeri
The University of Bristol
Bristol, UK
nirav.ajmeri@bristol.ac.uk

ABSTRACT

The evolving relationship between humans and technology entails increasing concerns about the impact on ethical issues such as bias, unfairness, and lack of accountability. There is thus a need for consistent responses to multiple-user social dilemmas that arise during interactions in sociotechnical systems, where combinations of humans and technical agents work together as ethical duos. The outcomes of these systems can be evaluated by the values that participating humans hold, which in turn influence the development of norms used to guide acceptable behaviour. However, when values are misjudged or norms conflict, dilemmas arise that must be resolved in satisfactory ways.

To examine these dilemmas, we adopt a macro ethics perspective where ethics is addressed via the governance of sociotechnical systems with multiple agents (rather than through the actions of single agents). We propose that to produce satisfactory outcomes, systematic methodologies be developed to consistently integrate normative ethical principles in reasoning capacities. The application of these ethical principles would enable practitioners to think analytically and systematically about the multiple-user social dilemmas that occur in these systems, in order to resolve them in satisfactory ways. To achieve this, we need to (1) categorize ethical principles not yet used in AI, form new ways of (2) systematically integrating ethical principles into reasoning, and use these new ways to (3) develop consistent responses to multiple-user social dilemmas.

KEYWORDS

Ethics in the large; values; norms; multi-user social dilemma

ACM Reference Format:
Jessica Woodgate and Nirav Ajmeri. 2022. Macro Ethics for Governing Equitable Sociotechnical Systems: Blue Sky Ideas Track. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 5 pages.

1 INTRODUCTION

Historically, much fairness literature has concentrated on binary classification algorithms (e.g. an algorithm that decides whether or not an applicant will be granted a bank loan). As highlighted by Murukannaiah and Singh [43], there is a need to look at AI ethics from the perspective of sociotechnical systems (STS), where humans and agents work together in ethical duos. This is because, Corbett-Davies and Goel [15] argue, statistical limitations entail that focusing on one algorithm does not properly address true fairness. Contextual factors (where context is understood as spheres of information, Nissenbaum [45] defines) must thus be accounted for, considering the needs of others as emphasized by Ajmeri et al. [6]. Therefore, Murukannaiah et al. [42] and Dubljević et al. [21] support that the ethical implications of MAS should be considered through the interaction of social (people and organisations) and technical tiers (computers and networks), as explained by Kafalı et al. [25], rather than focusing on the technical tier alone.

Within the domain of STS, there is also a difference in (1) ethics in the large or *macro ethics* and (2) ethics in the small or *micro ethics*, as explored by Chopra and Singh [12].

Micro ethics focuses on the more narrow perspective of a single agent, and the ethical implications of that agent's decision making within an STS. The issue with viewing ethics from a micro perspective is that agents do not exist in isolation, and it thus fails to encapsulate all of the relevant factors that go into ethical reasoning. Micro ethics is therefore too narrow to properly inform the design of technical agents with ideas from ethics.

Macro ethics on the other hand, focuses on making computational the governance of the STS in which agents are embedded, which is when participants attempt to align norms with their values. This perspective better addresses the full scope of ethically relevant features in multiple-user scenarios.

From the viewpoint of macro ethics, considering the role of values (what is important to us in life, Schwartz [48] explains) and norms (rules of expected behaviour, Morris-Martin et al. [40] define) is key to promoting equitable governance of STS, as supported by Ajmeri et al. [1]. Previous research into the broader concept of data governance such as Floridi [24] examines methods to improve data quality, yet does not incorporate the roles of values and norms. Research into value elicitation such as Liscio et al. [34] examines how context-specific values can be identified for ethical reasoning. However, important issues highlighted by Dignum and Dignum [19] must be addressed, relating to resolving the dilemmas that arise from values or norms conflicting.

As postulated by Dignum [18], it may be impossible to achieve perfect fairness in these dilemmas; what is fair for one group may be unfair for another. We thus suggest instead aiming for satisfactory outcomes that promote equitable systems (which meet the needs of different (groups of) stakeholders, as Perry [46] suggests, and involve the human aspect in the evaluation and creation of AI systems, as explained by Gilbert [22]). These systems have a higher goal of fairness. However, it is accepted that outcomes may not be perfectly fair or ethical. To obtain these satisfactory outcomes, we suggest that ethical principles should be implemented in reasoning.

Resolving these dilemmas has been targeted by Ajmeri [3], through the application of a single ethical principle. However, as argued by

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

Burton et al. [9], it is important to consider a variety of different approaches in ethical thinking. Challenges thus remain in creating satisfactory responses to multiple-user social dilemmas in which values or norms conflict. These responses should aim towards the development of equitable systems with the goal of fairness. This thus gives rise to an expansion of the scope of macro ethics and MAS to resolve multiple-user social dilemmas through the integration of a variety of ethical principles in reasoning.

2 SOCIOTECHNICAL SYSTEMS

The benefit of taking a holistic, sociotechnical standpoint is that it is a more accurate representation of the realities of ethical concerns. As Yazdanpanah et al. [55] and Chopra and Singh [12] explain, in many AI systems there is a social architecture that imposes regulations upon a technical architecture. For example, an AI agent that manages phone calls has a social architecture (interaction with the user) that regulates the technical architecture (the AI mechanisms) to ring or not ring the phone.

Therefore, Ajmeri and Murukannaiah [5] and Murukannaiah et al. [42] argue that we should first understand what ethics means at the social tier, and then bring that understanding to the technical tier so that it reflects ethical principles, rather than targeting the technical tier in isolation. It is important to shift ethical analysis from looking at the algorithm or technical aspect alone to the social perspective because, as stated by Etzioni and Etzioni [23], humans are the ones who have the basic attributes needed for moral agency. And, it is the humans, who have goals and values, and who are socially related to other humans.

The amalgamation of social and technical tiers is encapsulated through the concept of STS. Adapted from Murukannaiah and Singh [43], the STS is constituted of many human-agent duos, who in turn are encompassed by the STS. The non-discrimination of a group of duos corresponds to the (group) fairness of the STS. The values of the human guide the human-agent duo, and their reasoning is informed by the STS. Their decisions are then verified with respect to the norms of the STS, which operationalize the collective values of the participant duos. Likewise, the human-agent duos influence the governance of the STS (how the STS is administered by its participants, as defined by Singh [52]) by promoting their values and aims, and the STS is validated to the extent that it aligns to the values of its participants. Values and norms reinforce each other as norms emerge from values, and established norms then influence an individual in developing value preferences.

Governing Equitable Systems

Under the macro ethics perspective, norms and values are a crucial part in governing STS in equitable ways, as supported by Ajmeri et al. [3]. Considering values is key to ensuring satisfactory outcomes, Dubljević et al. [20] argue, and incorporating norms helps to regulate behaviour and ensure it is consistent with human values, as Montes and Sierra [38], Kafalı et al. [26], and Tzeng et al. [54] suggest. A typical example of where this would be necessary, as highlighted by Sen and Airiau [50], is in resolving social dilemmas where participants have different preferences.

However, there are issues that arise with the role of norms in social dilemmas. Ajmeri et al. [3] highlight how actions that comply with social norms are deemed to be legitimate, but legitimate actions may not be just or ethically appropriate. There are some situations where it might actually be better to violate norms, as stated by Morris-Martin et al. [40]. There thus needs to be some way of systematically making decisions about whether a norm should be violated or complied with; Yazdanpanah et al. [56] advocate for the development of norm ranking tools. This demonstrates the need to evaluate norms in equitable ways, and decide which ones should be followed and which should be violated.

Previous works such as Ajmeri et al. [3] and Kayal et al. [27] resolve norm conflicts by understanding the individual value preferences of multiple users over the relevant norms. Research into preference elicitation such as Braziunas and Boutilier [8] and Le Dantec et al. [31] could be utilised to gain these value preferences. Once the preferences have been elicited, works such as Mosca and Such [41] and Kurtan and Yolum [30] exemplify how preferences could be used in decision making. However, there are issues that arise when stakeholders have different value preferences.

Thinking about values is challenging for humans, Liscio et al. [34] suggest, and there are thus fundamental questions associated with how to prioritize them, Burton et al. [9] argue. Under macro ethics of STS, stakeholders govern by trying to align the norms with their values yet, as highlighted by Dignum [18], values may conflict. To govern equitable systems, we need to form ways to systematically guide and assist AI in reasoning about values, and decide which ones AI should elicit, learn, or align with, Etzioni and Etzioni [23] convey.

Multiple-user social dilemmas are thus scenarios in which a predicament arises that impacts multiple stakeholders (people affected by the outcome, Ajmeri et al. [2] explain). In these social dilemmas, there are cases where multiple norms may conflict with each other, one or more norms conflict with the value preferences of a user, or value preferences of one user conflict with those of other users. These dilemmas do not have to occur in extreme trolley problem cases as previous research has largely focused on, Etzioni and Etzioni [23] suggest, but also take place in mundane scenarios. For example, an app that helps a group of friends to decide where to eat must take into account relevant contextual factors, including values and norms, in order to find a satisfactory outcome. Under macro ethics of STS the friends govern by attempting to align the norms and decisions that are made with their values; however, dilemmas may arise. To enable equitable governance of STS, these dilemmas must be resolved in satisfactory ways. Thus, Murukannaiah et al. [42] argue that STS should provide social and technical controls to resolve norm and value conflicts in satisfactory ways. The consideration of ethical principles may aid the development of these social and technical controls, as they help practitioners to think analytically about the needs of the STS in question.

Ethical principles are understood here as operationalizable rules inferred from philosophical theories (e.g. Utilitarianism), which guide decision-makers in making normative judgments and determining the moral permissibility of concrete courses of actions according to McLaren [37] and Lindner et al. [33].

The abstraction of ethical principles, Dennis et al. [16] explain, allows for their applicability in a wide range of situations. They can thus facilitate choosing the norms and values that social and technical controls of the STS should align with, by determining the

moral permissibility of different courses of actions. This will promote equitable systems where the application of ethical principles can help to resolve dilemma scenarios in satisfactory ways.

3 OPEN CHALLENGES AND OPPORTUNITIES

A challenge from the macro ethics perspective of STS is how to create social and technical controls that work towards satisfactory outcomes in multiple-user settings where dilemmas arise concerning values and norms. Open research objectives encompass addressing these problems through the application of ethical principles. These principles would help ascertain how values and norms should be identified and prioritized; those that align with a particular principle would be prioritized higher than those which do not.

Under this broad domain, we suggest that there are three key areas that need addressing, as shown in Figure 1. We advocate for the development of (1) a taxonomy of ethical principles previously under-utilized in AI and Computer Science that can be applied to multiple-user social dilemmas, which should then be adapted to create (2) systematic methodologies for applying ethical principles in STS, and argue that these methodologies can be applied to achieve (3) consistent responses to multiple-user social dilemmas.

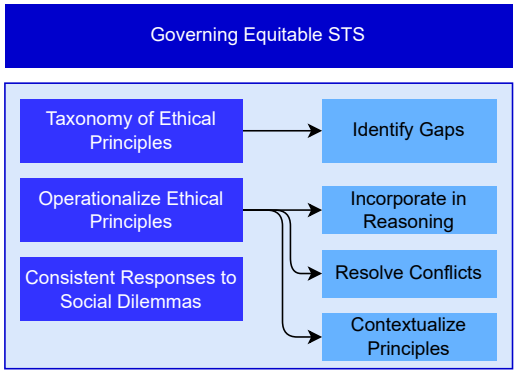

Figure 1: Open Challenges in Governing Equitable STS.

3.1 Taxonomy of Ethical Principles for AI

Q_1 Existing Gaps. What ethical principles exist that require further research in relation to AI ethics?

Motivation. Ethical principles can help to resolve multiple-user social dilemmas as they can guide normative judgments, understand different perspectives, and determine the moral permissibility of concrete courses of actions, as McLaren [37], Saltz et al. [47], and Lindner et al. [33] argue. Kim et al. [28] suggest that this is because ethical principles imply certain logical propositions that must be true in order for a given action plan to be ethical. Therefore, methodologies that an AI agent can use to incorporate ethical principles into making moral decisions in a wide variety of contexts may be useful in order to systematically think through dilemmas and promote satisfactory outcomes, Conitzer et al. [14] support. To enable the integration of ethical principles into reasoning capacities, the development of a comprehensive taxonomy of ethical principles would be beneficial. As Burton et al. [9] argue, ethical thinking should be fostered through the appreciation of a variety of different approaches, considering the strengths and limitations of each. Therefore, the incorporation of principles outside of the standard doctrine will improve the amplitude of ethical reasoning.

Challenge. A challenge therefore arises in looking outside of the domain of ethics typically used in AI or Computer Science, to examine a wider array of ethical principles. This may include principles that have been mentioned but not widely researched in Computer Science, or principles that exist in the domain of normative ethics but have never been utilised in Computer Science. Tolmeijer et al. [53] study how ethical principles relate to machine ethics, Yu et al. [57] create a taxonomy of ethical decision frameworks, and Dignum [17] gives a summary of normative ethics. However, to enable broader applicability, these works may benefit from a taxonomy including other ethical principles not yet widely researched.

Direction for MAS. Future research could include examining existing gaps to incorporate a wider array of ethical principles. We also emphasize the importance of researching principles from other cultures outside of the Western doctrine, which may aid better application to groups of stakeholders from diverse backgrounds. This could help form the groundwork for ethical decision support in macro ethics of STS.

3.2 Using Ethical Principles in Reasoning

Q_{2a} Integrating Principles into Ethical Reasoning. How can ethical principles be integrated into the ethical reasoning needed to govern STS?

Motivation. The incorporation of ethical principles in reasoning provides comprehensive guidance for decision making and can be utilized to analyze complex issues in depth, Canca [10] suggests. As Cheng et al. highlight [11], applying principles thus helps to put ethical AI into practice. This can be done, according to Cointe et al. [13], by incorporating principles into the reasoning capacities of agents. Therefore, ethical principles can help to provide guidance for ethical decision making about complex issues by enabling reasoning about the permissibility of certain actions.

Challenge. We thus need consistent methodologies to integrate ethical principles into the reasoning capacities used for equitable governance of STSs. This could be aided by applying previous research such as Loreggia et al. [35], related to evaluating if preferences are compatible with ethical principles, to the macro ethics of STS. These techniques should be used to discern the moral permissibility of certain outcomes, and the prioritization of values and norms according to ethical principles.

Direction for MAS. The development of ways in which the reasoning techniques used to govern STS can methodically incorporate ethical principles would be beneficial to ensure consistent responses in ascertaining whether a particular decision is acceptable.

Q_{2b} Conflicting Principles. How can we resolve cases in which different ethical principles result in different outcomes?

Motivation. There are difficulties associated with the application of normative ethics. Developing consistent ethical responses is a

much disputed issue; as Moor [39] explains, we have a limited understanding of what a proper ethical theory is. Therefore, people often disagree on the subject and individuals can have conflicting ethical intuitions and beliefs. According to McLaren [37], this means an ethics case may be resolved in multiple ways involving different possible actions and outcomes.

Challenge. A challenge is how to produce a perfectly fair or ethical outcome, as different principles may result in quite different scenarios. This challenge supports the idea that it is more achievable to aim for satisfactory outcomes. The application of ethical principles can guide practitioners towards such satisfactory outcomes. However, the heterogeneous nature of ethics means that the outcomes may not be perfectly fair or ethical. As explained by Sen [49], different ethical principles are not mutually compatible and may lead to different outcomes. There is thus a need to understand the distribution of how various ethical principles could be satisfied, and future work should study ways to resolve potential conflicts. We need to examine if it is preferable to (dis)satisfy all relevant principles equally or partially satisfy some and fully satisfy others.

Direction for MAS. Consistent methodologies for resolving the cases where ethical principles lead to different and potentially conflicting outcomes would thus benefit the ability to govern equitable STS. In doing this, the distribution of how different ethical principles are satisfied should be examined. New methodologies could leverage works such as Ajmeri et al. [4] on argumentation schemes to support decision making by capturing design rationale and evidence.

Q_{2c} Contextualising Principles. How can abstract ethical principles be applied to specific contexts?

Motivation. There are also difficulties with the application of principles, as highlighted by McLaren [37] and Madaio et al. [36], in the gap between abstract, open-textured rules and concrete facts. Abstract principles, they argue, contain open-textured terms and can be difficult to apply as they are subjective to interpretation and may have different meanings in different contexts. Although abstract principles are rules, they can't be applied without some way to "bridge the gap" between the rules' abstractions and concrete fact situations.

Challenge. Binns [7], Dennis et al. [16], and McLaren [37] therefore suggest that abstract principles need to be supplemented with contextual facts and empirical claims about how and why certain circumstances obtain. Whilst relevant facts (which may include observations of human values and preferences) do not by themselves decide what is ethical, they do factor into ethical assessment, Kim et al. [28] argue. To support this, Ajmeri et al. [1] suggest that context is crucial in determining courses of action, and a challenge therefore emerges in applying abstract principles to specific contexts.

Direction for MAS. Systematic ways of combining ethical principles with contextual information (such as values and norms) to reach satisfactory outcomes would aid the equitable governance of STS, by enabling the application of abstract principles to concrete facts. This may be aided by adapting work such as Kökciyan and Yolum [29] related to utilising context in reasoning.

3.3 Consistent Responses to Social Dilemmas

Q_3 Consistent Responses. How can we create consistent responses to multiple-user social dilemmas?

Motivation. The inevitability of dilemmas when viewing STS from the perspective of macro ethics, as supported by Dignum [18] and Morris-Martin et al. [40], entails the importance of developing consistent responses to them. As denoted by Conitzer et al. [14], in social dilemmas not everyone will agree about which factors are morally relevant, and even fewer people will agree about which factor is the most important. To create satisfactory outcomes, Leben [32] states that we need to develop consistent responses, despite the fact that stakeholders may disagree.

Challenge. Practitioners thus need clear guidance about how dilemmas in STS can be resolved consistently and in satisfactory ways, Shneiderman [51] denotes. One way we may be able to do this is by applying normative ethical theories to systematically analyse dilemmas, Saltz et al. [47] suggests. There thus exists a challenge in developing consistent responses to dilemma scenarios in which stakeholders may disagree. This does not mean responding to multiple-user social dilemmas in the same way each time. Instead, we advocate for the development of methodologies that aid the integration of ethical principles into reasoning capacities in reproducible ways.

Direction for MAS. One approach to address the challenge of consistently responding to social dilemmas in satisfactory ways is through the application of normative ethics, which could help to systematically analyze dilemmas. This could be achieved through the application of methodologies to integrate a variety of ethical principles into reasoning in reproducible ways.

4 CONCLUSION

To support the overarching goal of ethical MAS, we need to appreciate the interaction of both the social and technical tiers, encapsulated in the concept of sociotechnical systems. In examining STS, we should shift perspective from micro ethics (examining the actions of individual agents) to macro ethics, examining the governance of STS and considering how values and norms in ethical reasoning. Within these systems, multiple-user social dilemmas will inevitably arise in which norms or values conflict. To achieve satisfactory outcomes, we need consistent responses to these dilemmas. One approach to attain consistent responses is applying ethical principles to reason about existing values and norms and considering the implications for all stakeholders. Future research should address the development of a taxonomy of ethical principles for AI, including under-utilized principles and principles outside of the Western doctrine. We need to create systematic methodologies to integrate these principles into reasoning, resolve conflicts, and bridge the gap between their abstract nature and the concrete facts to which they must be applied.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful comments and suggestions. NA thanks Munindar P. Singh and Pradeep Murukannaiah for the discussions that form the basis of this paper.

REFERENCES

[1] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. 2018. Designing Ethical Personal Agents. *IEEE Internet Computing* 22, 2 (2018), 16–22.

[2] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. 2018. Robust Norm Emergence by Revealing and Reasoning about Context: Socially Intelligent Agents for Enhancing Privacy. In *Proc. IJCAI*. Stockholm, 28–34.

[3] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. 2020. Elessar: Ethics in Norm-Aware Agents. In *Proc. AAMAS*. Auckland, 16–24.

[4] Nirav Ajmeri, Chung-Wei Hang, Simon D. Parsons, and Munindar P. Singh. 2017. Aragorn: Eliciting and Maintaining Secure Service Policies. *IEEE Computer* 50, 12 (Dec. 2017), 50–58.

[5] Nirav Ajmeri and Pradeep Murukannaiah. 2021. Ethics in Socio-Technical Systems. In *Bristol Interactive AI Summer School*. The University of Bristol, Bristol.

[6] Nirav Ajmeri, Pradeep K. Murukannaiah, Hui Guo, and Munindar P. Singh. 2017. Arnor: Modeling Social Intelligence via Norms to Engineer Privacy-Aware Personal Agents. In *Proc. AAMAS*. São Paulo, 230–238.

[7] Reuben Binns. 2018. Fairness in Machine Learning: Lessons from Political Philosophy. In *Proc. 1st Conference on Fairness, Accountability and Transparency*. PMLR, New York, 149–159.

[8] Darius Braziunas and Craig Boutilier. 2006. Preference Elicitation and Generalized Additive Utility. In *Proc. AAAI* Boston, 1573–1576.

[9] Emanuelle Burton, Judy Goldsmith, Sven Koenig, Benjamin Kuipers, Nicholas Mattei, and Toby Walsh. 2017. Ethical Considerations in Artificial Intelligence Courses. *AI Magazine* 38, 2 (July 2017), 22–34.

[10] Cansu Canca. 2020. Operationalizing AI Ethics Principles. *Communications of the ACM* 63, 12 (Nov. 2020), 18–21.

[11] Lu Cheng, Kush R. Varshney, and Huan Liu. 2021. Socially Responsible AI Algorithms: Issues, Purposes, and Challenges. *JAIR* 71 (August 2021), 1137–1181.

[12] Amit Chopra and Munindar Singh. 2018. Sociotechnical Systems and Ethics in the Large. In *Proc. AIES*. New Orleans, 48–53.

[13] Nicolas Cointe, Grégory Bonnet, and Olivier Boissier. 2016. Ethical Judgment of Agents' Behaviors in Multi-Agent Systems. In *Proc. AAMAS*. Singapore, 1106–1114.

[14] Vincent Conitzer, Walter Sinnott-Armstrong, Jana S. Borg, Yuan Deng, and Max Kramer. 2017. Moral Decision Making Frameworks for Artificial Intelligence. In *Proc. AAAI*. Honolulu, 4831–4835.

[15] Sam Corbett-Davies and Sharad Goel. 2018. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *CoRR* abs/1808.00023 (2018), 1–25. arXiv:1808.00023 http://arxiv.org/abs/1808.00023

[16] Louise Dennis, Michael Fisher, Marija Slavkovik, and Matt Webster. 2016. Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems* 77 (2016), 1–14.

[17] Virginia Dignum. 2019. *Ethical Decision-Making*. Springer, Cham, 35–46.

[18] Virginia Dignum. 2021. The Myth of Complete AI-Fairness. In *Artificial Intelligence in Medicine*, Allan Tucker, Pedro Henriques Abreu, Jaime Cardoso, Pedro Pereira Rodrigues, and David Riaño (Eds.). Springer, Cham, 3–8.

[19] Virginia Dignum and Frank Dignum. 2020. Agents Are Dead. Long Live Agents!. In *Proc. AAMAS*. Auckland, 1701–1705.

[20] Veljko Dubljević, Sean Douglas, Jovan Milojevich, Nirav Ajmeri, William A. Bauer, George F. List, and Munindar P. Singh. 2021. Moral and Social Ramifications of Autonomous Vehicles. *CoRR* abs/2101.11775 (Jan. 2021), 1–8. arXiv:2101.11775

[21] Veljko Dubljević, George F. List, Jovan Milojevich, Nirav Ajmeri, William Bauer, Munindar P. Singh, Eleni Bardaka, Thomas Birkland, Charles Edwards, Roger Mayer, Ioan Muntean, Thomas Powers, Hesham Rakha, Vance Ricks, and M. Shoaib Samandar. 2021. Toward a Rational and Ethical Sociotechnical System of Autonomous Vehicles: A Novel Application of Multi-Criteria Decision Analysis. *PloS ONE* 16, 8 (Aug. 2021), e0256224.

[22] Juan E. Gilbert. 2021. Keynote: EquitableAI. In *Proc.* . ACM, Yokohama, 1–2.

[23] Amitai Etzioni and Oren Etzioni. 2017. Incorporating Ethics into Artificial Intelligence. *The Journal of Ethics* 21 (Dec. 2017), 403–418. Issue 4.

[24] Luciano Floridi. 2018. Soft Ethics and the Governance of the Digital. *Philosophy & Technology* 31 (2018), 1–8. Issue 1.

[25] Özgür Kafalı, Nirav Ajmeri, and Munindar P. Singh. 2016. Revani: Revision and Verification of Normative Specifications for Privacy. *IEEE Intelligent Systems* 31, 5 (Sep 2016), 8–15.

[26] Özgür Kafalı, Nirav Ajmeri, and Munindar P. Singh. 2020. Specification of Sociotechnical Systems via Patterns of Regulation and Control. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 29, 1 (Feb. 2020), 7:1–7:50.

[27] Alex Kayal, Willem-Paul Brinkman, Mark A. Neerincx, and M. Birna van Riemsdijk. 2018. Automatic Resolution of Normative Conflicts in Supportive Technology based on user values. *ACM Transactions on Internet Technology (TOIT)* 18, 4, Article 41 (May 2018), 21 pages.

[28] Tae Wan Kim, John Hooker, and Thomas Donaldson. 2020. Taking Principles Seriously: A Hybrid Approach to Value Alignment. arXiv:2012.11705 [cs.AI]

[29] Nadin Kökciyan and Pınar Yolum. 2020. TURP: Managing Trust for Regulating Privacy in Internet of Things. *IEEE Internet Computing* 24, 6 (2020), 9–16.

[30] A. Can Kurtan and Pınar Yolum. 2020. Assisting humans in privacy management: an agent-based approach. *Autonomous Agents and Multi-Agent Systems* 45 (Dec. 2020), 7–40. Issue 1.

[31] Christopher A. Le Dantec, Erika Shehan Poole, and Susan P. Wyche. 2009. Values as Lived Experience: Evolving Value Sensitive Design in Support of Value Discovery. In *Proc. CHI*. Boston, 1141–1150.

[32] Derek Leben. 2020. Normative Principles for Evaluating Fairness in Machine Learning. In *Proc. AIES*. New York, 86–92.

[33] Felix Lindner, Robert Mattmüller, and Bernhard Nebel. 2019. Moral Permissibility of Action Plans. In *Proc. AAAI*. Honolulu, 7635–7642.

[34] Enrico Liscio, Michiel van der Meer, Luciano C. Siebert, Catholijn Jonker, Niek Mouter, and Pradeep K. Murukannaiah. 2021. Axies: Identifying and Evaluating Context-Specific Values. In *Proc. AAMAS*. London, 799–808.

[35] Andrea Loreggia, Nicholas Mattei, Francesca Rossi, and K. Brent Venable. 2018. Preferences and Ethical Principles in Decision Making. In *Proc. AIES*. New Orleans, 222.

[36] Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proc. CHI*. Honolulu, 1–14.

[37] Bruce M. McLaren. 2003. Extensionally defining principles and cases in ethics: An AI model. *Artificial Intelligence* 150, 1 (2003), 145–181.

[38] Nieves Montes and Carles Sierra. 2021. Value-Guided Synthesis of Parametric Normative Systems. In *Proc. AAMAS*. London, 907–915.

[39] James Moor. 2006. The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems* 21 (08 2006), 18–21.

[40] Andreasa Morris-Martin, Marina De Vos, and Julian Padget. 2019. Norm Emergence in Multiagent Systems: A Viewpoint Paper. *Autonomous Agents and Multi-Agent Systems (JAAMAS)* 33, 6 (2019), 706–749.

[41] Francesca Mosca and Jose M. Such. 2021. ELVIRA: An Explainable Agent for Value and Utility-Driven Multiuser Privacy. In *Proc. AAMAS*. London, 916–924.

[42] Pradeep K. Murukannaiah, Nirav Ajmeri, Catholijn M. Jonker, and Munindar P. Singh. 2020. New Foundations of Ethical Multiagent Systems. In *Proc. AAMAS*. Auckland, 1706–1710.

[43] Pradeep K. Murukannaiah and Munindar P. Singh. 2020. From Machine Ethics to Internet Ethics: Broadening the Horizon. *IEEE Internet Computing* 24, 3 (May 2020), 51–57.

[44] Luis G. Nardin, Tina Balke-Visser, Nirav Ajmeri, Anup K. Kalia, Jaime S. Sichman, and Munindar P. Singh. 2016. Classifying Sanctions and Designing a Conceptual Sanctioning Process Model for Socio-Technical Systems. *The Knowledge Engineering Review (KER)* 31 (March 2016), 142–166. Issue 02.

[45] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Washington Law Review* 79, 1 (Feb. 2004), 119–157.

[46] Laura Perry. 2009. Characteristics of equitable systems of education: A cross-national analysis. *European Education* 41, 1 (2009), 79–100.

[47] Jeffrey Saltz, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. 2019. Integrating Ethics within Machine Learning Courses. *ACM Transactions on Computing Education* 19, 4 (Aug. 2019), 1–26.

[48] Shalom H. Schwartz. 2012. An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture* 2, 1 (Dec. 2012), 3–20.

[49] Amartya Sen. 2011. *The Idea of Justice*. Belknap Press of Harvard University Press.

[50] Sandip Sen and Stéphane Airiau. 2007. Emergence of Norms Through Social Learning. In *Proc. IJCAI*. Hyderabad, 1507–1512.

[51] Ben Shneiderman. 2021. Responsible AI: Bridging from Ethics to Practice. *Communications of the ACM* 64, 8 (July 2021), 32–35.

[52] Munindar P. Singh. 2013. Norms As a Basis for Governing Sociotechnical Systems. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 1, Article 21 (Dec. 2013), 23 pages.

[53] Suzanne Tolmeijer, Markus Kneer, Cristina Sarasua, Markus Christen, and Abraham Bernstein. 2021. Implementations in Machine Ethics: A Survey. *ACM Computing Surveys* 38 (2021), 1–38.

[54] Sz-Ting Tzeng, Nirav Ajmeri, and Munindar P. Singh. 2021. Noe: Norms Emergence and Robustness Based on Emotions in Multiagent Systems. In *Preproceedings of the International Workshop on Coordination, Organizations, Institutions, Norms and Ethics for Governance of Multi-Agent Systems (COINE)*. London, 1–15.

[55] Vahid Yazdanpanah, Enrico Gerding, Sebastian Stein, Corina Cirstea, M.C. Schraefel, Timothy James Norman, and Nick Jennings. 2021. Different Forms of Responsibility in Multiagent Systems: Sociotechnical Characteristics and Requirements. *IEEE Internet Computing* 6, 25 (2021), 15–22.

[56] Vahid Yazdanpanah, Enrico H. Gerding, Sebastian Stein, Mehdi Dastani, Catholijn M. Jonker, Timothy J. Norman, and Sarvapali D. Ramchurn. 2021. Responsibility ascription in trustworthy autonomous systems. In *Embedding AI in Society*. North Carolina State University, Virtual, 1–5.

[57] Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, Victor R. Lesser, and Qiang Yang. 2018. Building Ethics into Artificial Intelligence. *Proc. IJCAI*. Stockholm, 5527–5533.