Special Issue: Computational Ethics and Accountability

CRISTINA BAROGLIO, University of Torino, Italy
OLIVIER BOISSIER, Univ. Lyon, IMT Mines, Saint-Etienne, CNRS, Laboratoire Hubert Curien UMR 5516, France
AXEL POLLERES, Vienna Univ. of Economics and Business & Complexity Science Hub Vienna, Austria

Editorial

Computational Ethics and Accountability are becoming topics of increasing societal impact; in particular, on the one hand, in the context of recent advances in AI and machine-learning techniques, people and organizations accept decisions made for them by machines, be they buy-sell decisions, pre-filtering of applications, deciding which content users are presented, which personal data are shared and used by third parties, up to automated driving. In each of these application scenarios, where algorithms and machines support or even replace human decisions, ethical issues may arise.

On the other hand, algorithms and machines can play the role of verifying and cross-checking compliance of human players in analyzing digital records of social interactions, for instance, in business transactions and processes, but also in following rules of conduct in online social interactions. Closing the circle, based on such checks, again automated decisions may be implied that involve ethical requirements, such as nondiscrimination and fairness.

Apart from infamous "trolley problems" (Edmonds 2013), where even philosophers struggle to judge what is the "right" decision and going either way has dramatic impacts, there are more subtle everyday decisions that we now either happily delegate to machines or that are digitally recorded, which may have ethical implications: handling of personal data has to follow strict legal regulations, especially in social networks and in re-sharing personal data with businesses, (social) norms should be followed also in domains where automated agents enter interactions that were typically executed by human actors only, and fair business practices should be ensured within business processes, compliant with regulations, laws, and best practices.

In all these areas, at the very least, we expect transparency and accountability from automated decision and decision support systems: that is, we should require these systems to be transparent about how they make decisions and knowing who is accountable for those decisions and their effects.

Many voices demand a more responsible technology and engineering approach, such as articulated in the Copenhagen Letter (Techfestival 2017), or recent initiatives to standardize value-based ethically compliant system design, such as IEEE's P7000 (for Ethical Life-Cycle Concerns Working Group (EMELC-WG) 2017) family of standards.

Authors' addresses: C. Baroglio, Università degli Studi di Torino, Dipartimento di Informatica, via Pessinetto 12, I10149, Torino (TO), Italy; email: cristina.baroglio@unito.it; O. Boissier, Olivier Boissier, Mines Saint-Etienne, 158 cours Fauriel, 42023 Saint-Etienne, France; email: Olivier.Boissier@emse.fr; A. Polleres, Vienna University of Economics and Business, Institut für Informationswirtschaft, Gebäude D2, 3. OG Welthandelsplatz 1, 1020 Wien, Austria; email: axel.polleres@wu.ac.at.

1533-5399/2018/04-ART40
https://doi.org/10.1145/3195835

However, while these efforts put humans (decision makers as well as engineers) in the center to address the above-mentioned challenges, the stance we take in the present special issue is that Intelligent Systems and AI *themselves* can help to enable such accountability and transparency, thus, acting as technologies and technology design principles to enable rather than endanger ethically compliant, accountable, and eventually *sustainable* computing. Multi-agent systems, Semantic Web and Agreement Technologies, and Value-Sensitive Design are just some of the research areas whose methods and results can fruitfully support business ethics and social responsibility.

In this special issue, you will find a collection of articles that aim to make computational advances by approaching these challenges from different angles:

Automatic Resolution of Normative Conflicts in Supportive Technology Based on User Values. In their article, Kayal et al. (2018) discuss automatic resolution of potentially conflicting norm-based social commitments for data sharing. The proposed algorithm models and takes into account user values to suggest preferred resolutions for such conflicts, backed up by an empirical study with 396 participants that revolves around location sharing within families. The model of resolving conflicts between commitments to share the location in certain temporal and spatial contexts is, then, set in the context of a personal value profile that takes different dimensions (friendship, privacy, safety, independence, responsibility) and their relative importance into account. In their user study, the authors collected users' preferences among these dimensions (recruited according to specific inclusion and exclusion criteria via a crowd-sourcing platform) and tested the effectiveness of conflict resolutions according to these preferences as perceived by the participants, when presented with particular conflict scenarios. The article presents a detailed, very insightful analysis of the results.

Preserving Privacy as Social Responsibility in Online Social Networks. The next article in this volume, by Kekulluoglu et al. (2018), highlights the privacy problems around re-sharing others' personal data in social networks, which at the moment is largely uncontrolled. They discuss an agreement-technologies-based approach to enable the discussion of privacy configurations of posts by all affected parties: in their work, the authors develop a reciprocity-based negotiation approach to reach such privacy agreements, combining privacy rules with utility functions, which they evaluate and illustrate in an agent-based simulation. The article, first, presents a user study to better understand and classify privacy concerns, highlighting several important aspects that affect concerns around sharing certain types of content (e.g., posts/pictures expressing political views or private vacation) or possibilities to explicitly exclude certain people from (re-)shares. Next, the work presents an OWL-based ontology to model these aspects and suggests SWRL-based rules to express agents' content privacy and re-sharing rules on top of these. As the publishers' rules may differ with the affected agents' privacy rules, the authors present and automate a threshold-based negotiation strategy that takes both utility and privacy of re-sharing into account. The strategy may possibly breach certain privacy rules, if acceptable within thresholds. This work is relevant, because it models how we decide on privacy everyday, based on context and utility on a case-by-case basis, rather than based on strict, static rules.

Measuring Moral Acceptability in e-Deliberation: A Practical Application of Ethics by Participation. The next article, by Verdiesen et al. (2018), is also concerned with consensus and agreement, but—rather than about privacy and re-sharing policies—in the context of crowd-sourcing participatory decision making by harnessing open, wiki-like methodologies that promise to scale such participatory decision making in real-time. The authors study a challenging topic: the development of new forms of digitally moderated democratic processes, aimed at filtering, or moderating, extreme opinions without curbing the right of free speech. They name their approach the "Ethics by Participation" approach for participatory deliberation. Their proposal finds realization in the Massive Open Online Deliberation (MOOD) environment, a participatory platform that supports

and structures debates and provides debate outcomes in a form that is suitable for use by policy makers—to make better decisions. Moreover, the use of MOOD at the G1000 citizen participation event, which was organized in July 2017 by the city of Rotterdam, is explained.

Enhanced Audit Strategies for Collaborative and Accountable Data Sharing in Social Networks. Issues concerning trust and regulation of data access are explained and faced by Bahri et al. (2018), who in their article hope for a transition from the currently widely adopted centralized model of identity and data management to the adoption of a trust-less model, where Web users own and control their personal data. Instead of entrusting their privacy to service providers, and instead of relying on cryptography-based and computationally costly solutions, the idea is to rely on accountability and transparency in an open, trusted sharing environment. Each participant will specify provisions of fine-grained access rights to their own data. Compliance of the social network nodes behavior to such provisions will be checked *a posteriori*, with the aim of identifying delinquencies. The proposal advances the state-of-art in that it improves the audit strategy by which nodes to be checked are selected. A reputation-based strategy is proposed and is tested against a data-set collected from a real open social network. The strategy is shown to improve the efficiency of the system, in detecting bad behavior, by more than 50% with respect to previous experiments described in the literature.

easIE: Easy-to-Use Information Extraction for Constructing CSR Databases from the Web. Social accountability of companies is the issue faced by Gkatziaki et al. (2018). Environmental, social, and governance (CSR) performance are the aspects that are specifically considered by the authors of the article. Data that depicts a company as (un)ethical or (un)sustainable can have an impact on how that company is perceived and valued. The issue is that companies control the data upon which such evaluations are computed. To provide objective evaluations, a careful process is carried out by projects like WikiRate. The authors contribute to the computation of objective evaluations with a tool that is targeted at extracting data about companies from Web-accessible, semi-structured resources in a way that fits the justly tight requirements of the platform. The tool is not the only contribution of this work, which proposes also a data model for representing CSR information concerning companies and a data-set containing CSR data about 49,000 companies that were extracted by means of the tool.

Accountable Protocols in Abductive Logic Programming. In the last article of this special issue, Gavanelli et al. (2018) tackle the issue of accountability in a formal setting, which focuses on *S*CIFF, a language that was originally devised to formalize interaction protocols among agents, centered around the notion of expectation. This is done to better tackle the case of protocol violation, by triggering a process through which an agent can explain its behavior and, so, account for its actions and decisions. If an explanation that is consistent with its expected behavior is found, then the agent will not be considered as responsible for the state under scrutiny. The notion of accountability that is proposed is strictly related to that of responsibility for violation, and an interaction protocol will be accountable if and only if it will be proved to enable the identification of the responsible agents for all the possible violations. An interesting aspect of this formal work lies in its possible impact on the development of interaction protocols, as the definitions provide a guideline to the developer.

ACKNOWLEDGMENTS

We thank all the people who made this special issue possible for their support, that is, all authors, the anonymous reviewers, as well as in particular Prof. Munindar P. Singh, Editor in Chief, Transactions on Internet Technology.

REFERENCES

Leila Bahri, Barbara Carminati, Elena Ferrari, and Andrea Bianco. 2018. Enhanced audit strategies for collaborative and accountable data sharing in social networks. *ACM Trans. Internet Technol.* this volume (2018).

David Edmonds. 2013. *Would You Kill the Fat Man?: The Trolley Problem and What Your Answer Tells Us about Right and Wrong*. Princeton University Press, Princeton, NJ.

The Engineering Methodologies for Ethical Life-Cycle Concerns Working Group (EMELC-WG). 2017. Model Process for Addressing Ethical Concerns During System Design. Retrieved from https://standards.ieee.org/develop/project/7000.html.

Marco Gavanelli, Marco Alberti, and Evelina Lamma. 2018. Accountable protocols in abductive logic programming. *ACM Trans. Internet Technol.* this volume (2018).

Vasiliki Gkatziaki, Symeon Papadopoulos, Richard Mills, Sotiris Diplaris, Ioannis Tsampoulatidis, and Ioannis Kompatsiaris. 2018. easIE: Easy-to-use information extraction for constructing CSR databases from the web. *ACM Trans. Internet Technol.* this volume (2018).

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 Trans. Internet Technol.* this volume (2018).

Dilara Kekulluoglu, Nadin Kokciyan, and Pınar Yolum. 2018. Preserving privacy as social responsibility in online social networks. *ACM Trans. Internet Technol.* this volume (2018).

The Techfestival. 2017. The Copenhagen Letter. Retrieved from https://copenhagenletter.org/.

Ilse Verdiesen, Virginia Dignum, and Jeroen van den Hoven. 2018. Measuring moral acceptability in e-deliberation: A practical application of ethics by participation. *ACM Trans. Internet Technol.* this volume (2018).