Advertisement

nature masterclasses

Workshops include Scientific Writing and Publishing, Grant Writing, and Peer Review

Find out more

Published: 20 June 2023

Introduction: Digital Technologies and Human Decision-Making

Sofia Bonicalzi, Mario De Caro & Benedetta Giovanola

Topoi 42, 793–797 (2023) | Cite this article

298 Accesses | 1 Altmetric | Metrics

Advertisement

nature masterclasses

Courses on grant writing, presentations, data and more

Find out more

In the last few decades, the growing and crosscutting influence of digital technologies has had an impact on many existing research areas and has led to the flourishing of new research areas too, such as the ethics of artificial intelligence (AI) and roboethics. Most of the related research efforts in these areas are devoted to analyzing the ethical dimension of information and communication technologies (Floridi 2014), and of AI and algorithms (Liao 2020; Mittelstadt et al. 2016; Tsamados et al. 2021), and to devising morals applicable to intelligent machines (Floridi and Sanders 2004; Bostrom, forthcoming).

In this general framework, growing attention has been paid to the ethical design of digital technologies, where AI systems are considered as both objects (i.e., tools for/made by humans) and potential subjects (i.e., moral agents and patients): proposals aimed at developing ethics by design or value-sensitive design approaches to digital technologies (van den Hoven, Vermaas and van de Poel 2015; Friedman, Hendry and Borning 2017; Umbrello 2020) have significantly increased in the last years, especially in the framework of the design of AI for the social good (Umbrello and van de Poel 2021).

At the same time, attention has increasingly been paid also to specific ethical issues linked to digital technologies, such as privacy (Nissenbaum 2004; Richards 2015); tracking, monitoring, and processing of users' data (Wolmarans and Voorhove 2022); the impact on personal autonomy and identity (Botes 2022); opacity (Bonicalzi 2022; Herlocker, Konstan and Riedl 2000); biases and unfairness (Buolamwini and Gebru 2018; Eubanks 2018; Noble 2018; O'Neil 2016; Zimmermann and C. Lee-Stronach, 2022); manipulation (Klenk and Hancock 2019); the potential threats to democracy and society as a whole (Christiano 2022; Risse 2023); and the prospective loss of jobs and unemployment (Ernst 2022), just to mention the most relevant ones.

A key underexplored question concerns how and to what extent digital technologies may foster or hinder individual and collective human decision-making. This is a crucial issue to the extent that digital technologies shape the reality we live in, and affect the pre-conditions of our (supposedly free) choices: the progress and use of machine learning algorithms, based on deep learning architectures, often lead to non-explainable algorithmic decision-making (Pasquale 2015), involve biases (Benjamin 2019), and tend to create filter bubbles (Pariser 2011) or echo chambers (Sunstein 2008), thereby affecting our epistemic agency (Coeckelbergh 2022), predetermining the conditions and restricting the range of our choices (Giovanola and Tiribelli 2022). Moreover, we are delegating a great deal of the choice process to digital technologies, often without even realizing it (Royakkers et al. 2018) but nonetheless contributing to the gradual erosion of our moral capacities.

This situation raises underexplored ethical concerns and questions about the possibilities and constraints of human decision-making, in a reality that is more and more shaped by artificial intelligence and algorithms in a wide array of application domains, including—among others—social media communication and information management (Bozdag 2013; Shapiro 2020; Hinman 2008), advertising and marketing (Hildebrandt 2008; Tufekci 2015), recruiting and employment (Kim 2017), university admissions (Simonite 2020), housing (Barocas and Selbst 2016), credit lending (Devill 2013; Lobosco 2013; Lee and Floridi 2020), criminal justice (Berk et al. 2018), policing (Ferguson 2017), and healthcare (Danks and London 2017; Robbins 2019; Giovanola and Tiribelli 2023; Migliorelli et al. 2023).

To sum up, digital technologies increasingly shape the reality we live in as well as the boundaries of our moral agency, and raise the question of what the properly "human" side of decision-making is. At the same time, digital technologies have a pervasive impact on many dimensions of decision-making, which are increasingly automatized, supported by or delegated to AI and algorithms. Both aspects have been intensively investigated in isolation. However, to date little effort has been made to explore their connections—i.e., to investigate how digital technologies foster or hinder decision-making, and to examine whether AI and algorithms affect the pre-conditions of our choices and moral agency.

This edited collection aims to contribute to filling these research gaps, bringing together philosophers and AI experts to analyze the interplay between digital technologies and human decision-making.

The articles included in the present collection offer an analysis of the existing and foreseeable ethical and social challenges that AI technologies posit to human decision-making. Several contributions focus on concerns linked with specific emerging digital tools and areas of application—including online nudging (**Schmauder, Karpus, Moll, Bahrami and Deroy**), mental healthcare (**Minerva and Giubilini**), and recommender systems (**Bonicalzi, De Caro and Giovanola**)–, or with their techno-social implications, such as manipulation (**Ienca**) and increase in datafication (**Lavazza and Farina**). Other papers aim at critically overviewing or developing ethical frameworks and guidelines (**Liao**; **Mhlambi and Tiribelli**; **Deroy**), as well as political responses (**Hoeksema**), to be deployed in practical contexts where the usage of AI-based technologies is, or will be, skyrocketing.

Christian Schmauder, Jurgis Karpus, Maximilian Moll, Bahador Bahrami and Ophelia Deroy explain how AI methods provide new avenues, e.g., through an increase in fine-tuned personalization, for the development of effective forms of nudging. Nudge is a public policy that may often facilitate or improve human life across different decision-making domains, spanning from healthcare to the protection of the environment. However, in their contribution, the authors carefully highlight that outsourcing the design of nudges to AI systems whose workings are not entirely known or explainable—the infamous *black box problem* associated with AI—might be socially undesirable or even pernicious, with associated problems of accountability. In particular, since nudges notoriously exploit human weaknesses and fallacies, the obscure nature of AI-based nudge techniques may imply that users and programmers be unaware of the human cognitive processes that are selectively involved in reaching specific decision-making targets. On these grounds, the authors advocate for an interdisciplinary monitoring of the AI systems designing nudges.

In their contribution, **Francesca Minerva and Alberto Giubilini** consider whether AI is to be thought of as the future of mental healthcare. On the one hand, the possibility of deploying AI-based tools has considerably helped the administration and accessibility of healthcare worldwide. This appears to be particularly important in an age when mental health is globally deteriorating, demands for help are progressively augmenting, and AI might be able to outperform, or at least proficiently assist, human doctors throughout several medical procedures. Nonetheless, switching to forms of AI-based healthcare has some significant drawbacks, including, for instance, the dehumanization of the medical procedures. Thus, Minerva and Giubilini carefully address the pros of cons of using AI in the delivery of mental healthcare.

In their paper, **Sofia Bonicalzi, Mario De Caro and Benedetta Giovanola** contribute to the philosophical discussion concerning the ethical issues that have been raised in relation to the widespread diffusion of recommender systems throughout various life activities. In particular, they target the impact that recommender systems may have for what they term *descriptive autonomy*, a multi-faceted notion including both the potential to express oneself through action and the capacity to exert reasons-responsiveness and engage in reflective practices. Advocating for an ethically oriented implementation of recommender systems, but without indulging in the unrealistic defense of the *status quo*, the authors further articulate this challenge in terms of the risks of manipulation and deception associated with recommender systems, of their power to affect users' personal identity, and of their impact on knowledge acquisition and sharing processes as well as on critical thinking.

The focus on users' manipulation as a byproduct of the new forms of engagement brought about by the development of AI systems is central also to **Marcello Ienca**'s contribution. Here, the author provides an insightful critical overview of the literature on manipulation and AI technologies. Throughout the piece, Ienca highlights how manipulation is not uniquely associated with AI and its cognate tools. Indeed, AI-mediated forms of manipulation are not to be represented as qualitatively different from analogous dynamics occurring in human–human interactions. At the same time, they have unprecedented potential in terms of their capacity to target and steer people's decision-making, through their aptness to bypass users' cognitive defenses. On these grounds, the author discusses how various social actors, such as researchers, practitioners, and policymakers could deal with such challenges.

In our increasingly digitalized society, recommender systems (Bonicalzi, De Caro and Giovanola) and online nudging (Schmauder, Karpus, Moll, Bahrami and Deroy) represent just one of the many technological artifacts and tools with which users are learning to interact, and that are reshaping, to a large extent, their experience across different social environments. In their contribution, **Andrea Lavazza and Mirko Farina** provide a comprehensive overview of the ethical and social qualms associated with the intensive datafication that underlies the implementation of intelligent and sophisticated bio-technological unions. Their extensive analysis is complemented by a reflection on the desirability of the profound changes that such datafication processes will bring about. The discussion takes as a basis four fundamental insights regarding the impact that AI-based technologies may have on the experience of users and, more generally, citizens and workers. These respectively concern the tendency to erode human privacy, which may expand into forms of worrisome social and political control, the reduction of workers' freedoms, the limitations imposed on human creativity and imagination, and the overemphasis on efficiency and instrumental reason.

AI-based technologies deployed in sensitive social contexts, such as healthcare (see also Minerva and Giubilini), are under pressure in terms of the development of appropriate ethical frameworks and guidelines. In recent years, various social actors, ranging from private subjects to governmental agencies and institutions, have contributed to producing such documents. In reviewing them, **S. Matthew Liao** aims to go beyond the existing formalizations, emphasizing their limitations and flaws. While these normative tools share some fundamental principles, including the focus on autonomy and non-maleficence, they tend to be too abstract to have a profound impact on digitally advanced healthcare sectors and address the multiple concerns raised by AI-based technologies. Furthermore, at a more theoretical level, they often do not manage to appropriately justify the very same principles that they defend. To bridge these practical and theoretical gaps, Liao proposes an ethical framework that finds its proper justification in human rights theory, which is aptly extended to the healthcare domain and holds that people have rights to "the fundamental conditions for pursuing a good life".

The focus on the limitations of the current ethical frameworks for AI is central to **Sábëlo Mhlambi and Simona Tiribelli**'s contribution as well. Referring to the existing ethical frameworks, the authors provocatively call into question the almost exclusive emphasis on a liberal notion of autonomy as self-determination (see also Bonicalzi, De Caro and Giovanola). This notion is criticized as inadequate to account for the many senses in which human autonomy can be violated in the context of artificial decision-making. Furthermore, as this narrow notion of autonomy is grounded in Western traditional philosophy and linked with a history of colonization, the corresponding ethical frameworks may fail to understand the extent to which AI-related harms may cause substantial trouble to those who are already globally marginalized and disenfranchised. In the attempt to respond to this cultural challenge, Mhlambi and Tiribelli stress the need for a relational turn, rooted in moral philosophy and Ubuntu ethics, in the ethical frameworks regulating AI.

Another controversial notion associated with AI technologies and consistently deployed by ethicists, governmental agencies, and institutions is the label "trustworthy AI" In her contribution, **Ophelia Deroy** warns against the indiscriminate usage of ambiguous or controversial expressions, or *loose talk*, attributing human-like features to AI. By browsing the field of AI ethics and science communication, the author reviews the reasons and speaks out against the fragile justifications—ontological, legal, communicative, and psychological—that contribute explaining these questionable linguistic practices. In particular, in pointing at the potentially negative social consequences that this attitude may have, Deroy focuses on two problematic arguments that may underlie the tendency to anthropomorphize AI, i.e., the claim that discourses on the ethics of AI do not fundamentally require philosophical clarification, and the claim that such humanizing language appropriately matches how non-experts conceptualize AI.

While most of the papers hosted in this collection are concerned with the ethical challenges posited by digital technologies, **Bernd Hoeksema**'s contribution aims to articulate a political perspective, under the umbrella of republicanism, on online jerkish speech. Jerkish speech is here identified as the speech with which users show disregard for the perspective of others, notably when the latter are perceived as having a lower social status. While online jerkish speech might be wrongly considered scarcely impactful in comparison with forms of more explicit hate speech or online harms, it could lead to systemic or structural forms of domination or forms of micro-domination, the latter being individually inconsequential but nonetheless problematic when they are thought of in aggregate. In the paper, Hoeksema discusses how the republicanism program, having a focus on the notion of "domination", has the tools to explain why online jerkish speech is problematic as well as to develop an appropriate strategy to reduce its social impact.

Thus, as this brief overview aims to illustrate, this collection on *Digital Technologies and Human Decision-Making* aims to raise a discussion about the most urgent ethical and social challenges regarding the impact digital technologies (may) have on human decision-making, both for individuals and for society as a whole.

References

Barocas S, Selbst AD (2016) Big data's disparate impact. SSRN Electron J. https://doi.org/10.2139/ssrn.2477899

Benjamin R (2019) Race after technology: abolitionist tools for the new jim code. Polity, Medford

Berk R, Heidari H, Jabbari S, Kearns M, Roth A (2018) Fairness in criminal justice risk assessments: the state of the art. Sociol Methods Res. https://doi.org/10.1177/0049124118782533

Bonicalzi S (2022) A matter of justice. The opacity of algorithmic decision-making and the trade-off between uniformity and discretion in legal applications of artificial intelligence. Teoria 42(2):131–147. https://www.rivistateoria.eu/index.php/teoria/article/view/161

Botes M (2022) Autonomy and the social dilemma of online manipulative behavior. AI Ethics. https://doi.org/10.1007/s43681-022-00157-5

Bozdag E (2013) Bias in algorithmic filtering and personalization. Ethics Inf Technol 15:209–227. https://doi.org/10.1007/s10676-013-9321-6

Buolamwini J, Gebru T (2018) Gender shades: intersectional accuracy disparities. commercial gender classification. Proceedings of the 1st conference on fairness, accountability and transparency, PMLR, 81, 77–91

Christiano T (2022) Algorithms, manipulation, and democracy. Can J Philos 52(1):109–124

Coeckelbergh M (2022) The Political Philosophy of AI. Polity Press, Cambridge, UK

Danks D, London AJ (2017) Algorithmic bias in autonomous systems. Proceedings of the twenty-sixth international joint conference on artificial intelligence. International joint conferences on artificial intelligence organization, 4691–4697. https://doi.org/10.24963/ijcai.2017/654

Deville J (2013) Leaky data: how Wonga makes lending decisions. Consumer Market Studies, Charisma

Ernst E (2022) The AI trilemma: saving the planet without ruining our jobs. Front Artif Intell. 5:886561. https://doi.org/10.3389/frai.2022.886561

Eubanks V (2018) Automating inequality. How high-tech tools profile, police, and punish the poor. St Martin's Publishing, New York

Ferguson AG (2017) The rise of big data policing. Surveillance, race, and the future of law enforcement. New York University Press, New York

Floridi L (2014) The Fourth Revolution: how the Infosphere is Reshaping Human Reality. Oxford University Press

Floridi L, Sanders JW (2004) On the morality of artificial agents. Mind Mach 14(3):349–379

Friedman B, Hendry DG, Borning A (2017) A survey of value sensitive design methods", foundations and trends®. In: human–computer interaction 11 2, pp 63–125. https://doi.org/10.1561/110000001

Giovanola B, Tiribelli S (2022) Weapons of moral construction? On the value of fairness in algorithmic decision-making. Ethics Inf Technol. https://doi.org/10.1007/s10676-022-09622-5

Giovanola B, Tiribelli S (2023) Beyond bias and discrimination. Redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. AI & Soc 38:549–563. https://doi.org/10.1007/s00146-022-01455-6

Herlocker JL, Konstan JA, Riedl J (2000) Explaining collaborative filtering recommendations. CSCW '00: proceedings of the 2000 ACM conference on computer supported cooperative work. 241–250

Hildebrandt M (2008) Defining profiling: a new type of knowledge? In: Hildebrandt M, Gutwirth S (eds) Profiling the European Citizen. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6914-7_2

Hinman LM (2008) Searching ethics: the role of search engines in the construction and distribution of knowledge. In: Spink A, Zimmer M (eds) Web search. Information science and knowledge management. Springer, Berlin, p 14. https://doi.org/10.1007/978-3-540-75829-7_5

Kim PT (2017) Data-driven discrimination at work. 58 Wm. & Mary L. Rev, 857 (3). Accessed 11 Mar 2021, from https://scholarship.law.wm.edu/wmlr/vol58/iss3/4

Klenk M, Hancock J (2019) Autonomy and online manipulation. Internet policy review

Lee MSA, Floridi L (2020) Algorithmic fairness in mortgage lending: from absolute conditions to relational trade-offs. Mind Mach. https://doi.org/10.1007/s11023-020-09529-4

Liao M (ed) (2020) Ethics of artificial intelligence. Oxford University Press, New York

Lobosco K (2013) Facebook friends could change your credit score. CNN Business

Migliorelli L, Tiribelli S, Cacciatore A, Giovanola B, Frontoni E, Moccia S (2023) Accountable deep-learning-based vision systems for preterm infant monitoring. Computer. https://doi.org/10.1007/s11245-023-09922-5

Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L (2016) The ethics of algorithms: mapping the debate. Big Data & Soc. https://doi.org/10.1177/2053951716679679

Nissenbaum H (2004) Privacy as contextual integrity. Wash Law Rev 79(1):119–158

Noble SU (2018) Algorithms of oppression: how search engines reinforce racism. New York University Press, New York

O'Neil C (2016) Weapons of math destruction: how big data increases inequality and threatens democracy. Crown, New York

Pariser E (2011) The filter bubble: what the internet is hiding from you. Penguin, New York

Pasquale F (2015) The black box society: the secret algorithms that control money and information. Harvard University Press, Cambridge

Richards N (2015) Intellectual privacy. Rethinking civil liberties in the digital age. Oxford University Press, New York

Risse M (2023) The political theory of the digital age. Where artificial intelligence might take us. Cambridge University Press, Cambridge

Robbins S (2019) A misdirected principle with a catch: explicability for AI. Mind Mach 29(4):495–514. https://doi.org/10.1007/s11023-019-09509-3

Royakkers L, Timmer J, Kool L, van Est R (2018) Societal and ethical issues of digitization. Ethics Inf Technol 20(2):127–142. https://doi.org/10.1007/s10676-018-9452-x

Shapiro S (2020) Algorithmic television in the age of large-scale customization. Telev New Media 21(6):658–663. https://doi.org/10.1177/1527476420919691

Simonite T (2020) Meet the secret algorithm that's keeping students out of college. Wired, San Francisco

Sunstein C (2008) Democracy and the Internet. In: van den Hoven J, Weckert J (eds) Information Technology and Moral Philosophy. Cambridge University Press, Cambridge, pp 93–110

Tsamados A, Aggarwal N, Cowls J, Morley J, Roberts H, Taddeo M, Floridi L (2021) The ethics of algorithms: key problems and solutions. AI & Soc. https://doi.org/10.1007/s00146-021-01154-8

Tufekci Z (2015) Algorithmic harms beyond Facebook and Google: emergent challenges of computational agency. Journal on Telecommunications and High Technology Law, 13(203). Accessed 11 Mar 2021 from https://ctlj.colorado.edu/wp-content/uploads/2015/08/Tufekci-final.pdf

Umbrello S (2020) Imaginative value sensitive design: using moral imagination theory to inform responsible technology design. Sci Eng Ethics 26(2):575–595

Umbrello S, van de Poel I (2021) Mapping value sensitive design onto AI for social good principles. AI Ethics 1(3):1–14. https://doi.org/10.1007/s43681-021-00038-3

Van den Hoven J, Vermaas PE, van de Poel I (2015) Handbook of ethics, values, and technological design. Sources, theory, values and application domains, Springer. ISBN: 978-94-007-6969-4

Wolmarans L, Voorhove A (2022) What makes personal data processing by social networking services permissible? Can J Philos 52(1):93–108

Zimmermann A, Lee-Stronach C (2022) Proceed with caution. Can J Philos 52(1):6–25

Download references

Funding

The three authors benefitted from the PRIN Grant 20175YZ855 from the Italian Ministry for Education, University and Research (Ministero dell'Istruzione, dell'Università e della Ricerca). Benedetta Giovanola also benefitted from the Jean Monnet Chair (Grant Agreement 101085372) EDIT – Ethics for Inclusive Digital Europe, co-funded by the European Union. Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Author information

Authors and Affiliations

Department of Philosophy, Communication and Performing Arts, Roma Tre University, Rome, Italy
Sofia Bonicalzi & Mario De Caro

CVBE Cognition, Value and Behavior, Ludwig-Maximilians-Universität München, Munich, Germany
Sofia Bonicalzi

Department of Philosophy, Tufts University, Medford, USA
Mario De Caro & Benedetta Giovanola

Department of Political Sciences, Communication, and International Relations, University of Macerata, Macerata, Italy
Benedetta Giovanola

Corresponding author

Correspondence to Sofia Bonicalzi.

Additional information

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rights and permissions

Reprints and Permissions

About this article

Check for updates

Cite this article

Bonicalzi, S., De Caro, M. & Giovanola, B. Introduction: Digital Technologies and Human Decision-Making. *Topoi* **42**, 793–797 (2023). https://doi.org/10.1007/s11245-023-09939-w

Download citation

Published	Issue Date
20 June 2023	July 2023

DOI
https://doi.org/10.1007/s11245-023-09939-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Provided by the Springer Nature SharedIt content-sharing initiative

82.57.177.85
Not affiliated

Springer Nature
© 2023 Springer Nature