

Integrated Science

Volume 40

Editor-in-Chief

Nima Rezaei , Tehran University of Medical Sciences
Tehran, Iran

The **Integrated Science** Series aims to publish the most relevant and novel research in all areas of Formal Sciences, Physical and Chemical Sciences, Biological Sciences, Medical Sciences, and Social Sciences. We are especially focused on the research involving the integration of two or more academic fields offering an innovative view, which is one of the main focuses of Universal Scientific Education and Research Network (USERN), science without borders.

Integrated Science is committed to upholding the integrity of the scientific record and will follow the Committee on Publication Ethics (COPE) guidelines on how to deal with potential acts of misconduct and correcting the literature.


Mariafilomena Anzalone
Stefania Achella • Fiorella Battaglia
Anna Donise
Editors


Reconfiguring Human Autonomy

Conceptual Challenges and Ethical
Implications in the Age of AI


 Springer

Editors

Mariafilomena Anzalone 
Dipartimento per l'Innovazione Umanistica,
Scientifica e Sociale (DIUSS)
University of Basilicata
Potenza, Italy

Stefania Achella 
Dipartimento di Scienze Filosofiche,
Pedagogiche e Sociali
University of Chieti-Pescara
Chieti, Italy

Fiorella Battaglia 
Dipartimento di Studi Umanistici
University of Salento
Lecce, Italy

Anna Donise 
Dipartimento di Studi Umanistici
University of Naples Federico II
Naples, Italy

ISSN 2662-9461

ISSN 2662-947X (electronic)

Integrated Science

ISBN 978-3-032-14536-9

ISBN 978-3-032-14537-6 (eBook)

<https://doi.org/10.1007/978-3-032-14537-6>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Competing Interests

This publication was developed within the framework of the European Union – NextGenerationEU, under the National Recovery and Resilience Plan (PNRR), Mission 4, Component 2, PRIN 2022 PNRR Research Project “DigitHuman: Digital Transition and Human Values”. CUP: F53D23010740001.

Preface

The integration of algorithms and artificial intelligence into diverse domains—including education, healthcare, caregiving, and warfare—poses significant challenges to longstanding philosophical frameworks traditionally employed to conceptualize, interpret, and characterize human beings. Among these, the concept of autonomy can be identified as a central and enduring principle within the tradition of modern philosophy. Against the backdrop of human autonomy, which is traditionally understood as rational self-determination grounded in a free and morally responsible will, the emergence of so-called “autonomous” machine agents—capable of learning, decision-making, and action—challenges established assumptions and calls for renewed theoretical and ethical reflection. In response to the urgency posed by these radical transformations, the concept of autonomy has once again become a central focus of contemporary philosophical and ethical debate, particularly within moral theory, where it plays a foundational role in articulating principles of responsibility, agency, and moral accountability. Over the centuries, the paradigm of personal autonomy has been articulated through various conceptual frameworks, including rational, moral, political, and relational interpretations. Although the concept of personal autonomy plays a central role in both moral and political philosophy, its precise definition and scope remain a matter of considerable debate. Several influential philosophers have developed and accepted analyses of autonomy closely aligned with the idea that an autonomous person is their own legislator—someone who governs themselves and determines their actions independently. This interpretation ranges from Immanuel Kant’s foundational anchoring of autonomy in the moral rationality of the subject, through Harry J. Frankfurt’s “self-government” perspective, to the feminist framework of relational autonomy advanced by Marina Oshana, Natalie Stoljar, and Catriona Mackenzie, which challenges the traditional notion of autonomy as simply being one’s own legislator. Autonomy has consistently played a pivotal role in shaping the conception of human identity. Concurrently, the concept has gradually expanded, leading to the attribution of autonomy to increasingly broader and more diverse categories of subjects. However, in contemporary discourse, the attribute of “autonomous” is increasingly applied to objects as well, particularly in the context of advanced technologies. Autonomous machines and artificial intelligence systems—capable of learning, making decisions, and acting without direct human control—exemplify this shift. This extension of autonomy beyond human agents raises

complex theoretical and ethical questions about the nature of agency, responsibility, and moral accountability in relation to non-human actors. With the design and construction of “autonomous” machines—systems to which we delegate tasks traditionally performed by humans, often involving moral responsibility—the notion that autonomy is an exclusive human attribute increasingly comes into question. Through the integration of artificial intelligence, machine learning, and advanced decision-making capabilities, these machines replicate human autonomous action in a manner that is unprecedented both in scope and complexity. This technological evolution raises profound ethical challenges, particularly regarding accountability, responsibility, and the delegation of moral agency to non-human actors. Philosophically, it calls into question longstanding assumptions about the uniqueness of human autonomy, prompting a re-examination of what it means to be an autonomous agent in a world increasingly shared with intelligent machines. The blurring boundaries between human and machine agency demand new theoretical frameworks that can address issues such as machine moral status, the limits of algorithmic decision-making, and the conditions under which autonomy can be ascribed beyond the human subject. We are confronted with a challenge whose full contours remain yet to be defined. In our view, addressing this challenge requires focusing on the connection between human vulnerability and the fundamentally relational nature of human autonomy—an aspect emphasized, for example, in the concept of empathic relationship.

This volume is driven by two central aims: to rethink human autonomy considering the rise of “machine” autonomies and, simultaneously, to critically examine the descriptive and normative boundaries attributed to these forms of autonomy. The idea is neither to reinforce yet another dichotomy of “us” versus “the machines” nor to endorse optimistic narratives that portray AI as a means for humanity’s moral enhancement. Instead, this volume calls for an inversion of perspective—inviting readers to explore how debates surrounding the autonomy of artificial systems can “refract” and illuminate the limits, possibilities, and transformative potential inherent in human autonomy, both at the individual and collective levels. To this end, in contrast to oppositional approaches that regard machine autonomy as a threat to human autonomy, this book adopts a co-evolutionary and dialogical approach. This text explores how new forms of autonomy in artificial and technological contexts contribute to a broader, more pluralistic, and contextualized understanding of human autonomy. Rather than viewing autonomy as a fixed, monolithic notion, the text considers it a historically contingent concept subject to ongoing refinement. The challenge before us is to address these theoretical, ethical, and political issues by critically questioning the traditional approach. Rather than evaluating the “machinic” of artificial autonomy solely through human-centric parameters, we investigate the evolving forms of autonomy with the aim of fundamentally reconsidering the concept of human autonomy itself.

Articulated into three sections, the volume navigates through conceptual genealogy, ethical considerations, and practical applications.

In the first part, “What Is Autonomy in the Age of AI?”, contributions undertake the essential intellectual task of disentangling the idea of “autonomy” from its

exclusive human-centric ties. In this section, the question of autonomy is also addressed from a historical-philosophical perspective, questioning its foundations and transformations in the light of comparisons with artificial intelligence. The contributions analyse the aporias of the modern concept of autonomy, explore feminist and relational perspectives, focus on the limits of traditional categories in describing the agency of machines, and propose a critical analysis of the mimetic and perception-based models that shape our relationship with artificial agents. Starting from Mariafilomena Anzalone's account of the imitative devices that inform AI (Chap. 1), through Stefania Achella's feminist reinterpretation (Chap. 2), to the delimitation of the criteria of agency and moral *patience* in Paul Formosa, Inês Hipólito, and Thomas Montefiore's contribution (Chap. 3), this section shows how autonomy, far from being a unitary concept, is articulated in a plurality of dimensions: cognitive, relational, normative, perceptual, and projective. In this case, AI acts as a "litmus test", forcing us to distinguish and circumscribe what, in our tradition, has been called autonomy—freedom, responsibility, self-legislation—from what machines actually achieve—learning, adaptation, and action without supervision.

The second part, "Human and Machinic Autonomy. Ethical Implications", focuses on the ethical and social repercussions of the interaction between human agents and autonomous machine systems. This interaction is investigated with particular attention to specific autonomous technological systems such as AVs, personal avatars, autonomous weapons, and recommendation systems. By applying theoretical perspectives such as that of relational autonomy or reflecting on the conditions and specificities of personal and moral autonomy, the contributions of this section concretely explore the transformative effects of such interaction. Hugo Cossette-Lefebvre and Natalie Stoljar show how *recommender systems* shape—and sometimes distort—users' preferences, reproducing social biases (Chap. 4); Mario De Caro reopens the question of the moral status of strong AI (Chap. 5), whereas Fiorella Battaglia's contribution focuses on how the concept of personal autonomy changes when it is extended to artificial intelligence systems, with a specific focus on the domain of defence (Chap. 6); Fabio Fossa criticizes techno-solutionist views on human autonomy "enhanced" by autonomous vehicles (Chap. 7); Paula Sweeney, finally, investigates how personal avatars and synthetic agents challenge the identity continuity of the subject and the ownership of actions (Chap. 8). As a result, the essays in this section also draw a map of the new ethical-political challenges that emerge from the attribution of autonomy to artificial devices: from widespread responsibility and algorithmic manipulation to the displacement of action and new configurations of power.

The third part, "Autonomy and AI in Medicine and Health Care. Challenges, Problems, and Opportunities", explores one of the most delicate and significant fields of AI application: the health sector. Here, the focus falls on the key areas of health and care, where the value of autonomy meets the fragility of bodily vulnerability. In this context, the introduction of AI is generating enormous expectations, both from the point of view of medical and pharmacological research and for clinical practice and the management of health activities. In this context, the question of

autonomy, as the essays in this section aptly show, must be addressed from several points of view—that of doctors, that of patients, and that of their relationship. Special focus is placed on the impact of predictive systems, diagnostic chatbots, and choice-supporting algorithmic tools on the decision-making autonomy of the patient and of the doctor, which redefines the terms and quality of their relationship. The authors’ analytical viewpoints—which encompass the ethics of care, narrative medicine, culturally attuned healthcare, and biopolitical theory—illustrate that the concept of autonomy has become a crucial intersection of epistemological, cultural, and political issues in contemporary discourse. Miranda Boldrini argues in favour of a relational view of autonomy that can enrich the ethics of AI in the health sector, focusing attention on the issues of trust, responsibility, and care (Chap. 9). While assessing risks and opportunities deriving from the impact of AI in health decision-making, Andrea Berber and Jelena Mijić identify the interest and well-being of the patient as the reference value apt to balance patient autonomy and that of the doctor (Chap. 10). Deivide Garcia da Silva Oliveira elaborates and presents a philosophical model for the doctor-patient relationship in the age of AI inspired by Feyerabend’s pluralism (Chap. 11). Chantal Marazia, Vasilija Rolfes, and Fabio De Sio focus on the concept of “algorithmic humility” to promote a relational and reflective approach to health care, in which patient autonomy is established as a dynamic and socially rooted process (Chap. 12). Finally, Chetan Shirvankar asks how to promote a human-centred care system, in the light of the growing datafication of the doctor-patient relationship, and the risks of epistemic injustice introduced by the anthropomorphization of machines in the clinical field (Chap. 13).

As seen from the individual contributions, the incorporation of AI in the most varied contexts of our private and public life already widely shows its enormous transformative potential. When systems are transparent and interactive, they can strengthen the perception of control and individual choice; in contrast, algorithmic opacity can precisely undermine the element of trust that supports autonomy. For this reason, the lived experiences of not only students, teachers, patients, and health professionals but also, more simply, all the various people who interact with AI for different reasons are crucial: determining when and why AI emancipates or makes people dependent means redefining autonomy on the basis of concrete relationships rather than abstract normative statements. It is precisely in this framework that the paradigm of *relational autonomy*, developed by feminist theory, can play a key role and is recalled in many of the contributions in this volume: overcoming an atomistic and rationalist approach requires recognizing the structural links and inequalities that AI risks amplifying. A path is therefore outlined to counter damage to the most vulnerable subjects and de-emphasize gender biases and algorithmic disparities. Moreover, the volume highlights the need to carefully circumscribe the analogy between human autonomy and machine autonomy. While AI systems can replicate some functions of autonomous decision and action, these functions do not erase the ontological and moral distance between a human agent—capable of self-reflection—and a system, however complex, still lacking that internal evaluative dimension. Even in the military, for example, delegating life and death decisions to autonomous systems is not equivalent to transferring moral autonomy, but it definitely introduces

fractional and controversial responsibilities. In the health and training field, the tension between responsibility and human autonomy is particularly strong. Systems such as recommenders or virtual avatars redefine the relationship between subject and knowledge as well as between representation and action. However, their effectiveness in increasing capacity for action and accessibility to information is accompanied by the risk of addiction, loss of authority, and agency crises. The figure of the doctor, for example, is not replaced by an algorithm but is certainly transformed by it, while the empowerment of the patient can oscillate between information consumption and passive delegation.

This highlights a fundamental regulatory issue. To support or rebuild autonomy in an algorithmically driven world, AI systems must be transparent, culturally inclusive, and designed on the basis of ethical criteria. This entails a standpoint well beyond reductionist or techno-solutionist approaches. It also implies acknowledging the fact that autonomy is not a static attribute, typical of an isolated individual, but a dynamic, distributed process, negotiated in contexts of care, education, and technical-social interaction.

This collection of contributions therefore converges on one underlying claim: human autonomy and machine autonomy should not be understood in terms of mere conflict or substitution but rather in their complex interrelation. Intelligent are neither simple neutral “tools” nor independent subjects; they reorient social practices, forms of interaction, and decision-making processes, thus redefining the very conditions of the possibility of autonomous action. Symmetrically, our moral categories—responsibility, accountability, agency—condition the design, regulation, and use of AI.

This volume finally invites us to overcome a dichotomously binary vision: autonomy is not a fixed attribute to be defended or delegated but rather a dynamic field of negotiation in which the recognition of subjects, the distribution of skills, and the construction of inclusive future scenarios play out. For this reason, we think that reconfiguring autonomy means, ultimately, reconfiguring what is human today. This amounts to accepting that our practical identity no longer coincides with an ideal of rational and individual self-sufficiency but is articulated and structured in the social bonds, in the technical environments, and in the socio-algorithmic networks that we inhabit.

The complex nature of this process is reflected in the essays collected in this book with their wide range of analyses and proposals, with their primary objective being the critical, not naive, exploration of a key question. Moving between moral philosophy, gender studies, phenomenology, the philosophy of technology, medical humanities, and ethical engineering, they show how the challenge of AI can become an opportunity to expand, rather than narrow, the horizons of individual and collective autonomy.

With this volume, we hope to contribute to a critical discussion that does not yield naive enthusiasm or forms of apocalyptic Luddism, which are rather common in contemporary debate. This discussion should allow the reader to see in artificial intelligence not only a technical tool or an entity “other” from us but also a mirror—sometimes faithful, sometimes deforming—of our most deeply rooted

categories of understanding and evaluation of reality. Precisely in this ambivalent duplicity of AI lies its philosophical richness. Here lies the opportunity to walk past AI as a mere object, the use of which should be regulated, and embrace the investigation of who we are and who we want to be in the world we are helping to build.

Potenza, Italy
Chieti, Italy
Lecce, Italy
Naples, Italy

Mariafilomena Anzalone
Stefania Achella
Fiorella Battaglia
Anna Donise

Acknowledgements

We are grateful to Tessa Marzotto Caotorta for her essential support in preparing this volume, including editing, formatting, reference verification, and coordinating communication with contributors.

Contents

Part I What Is Autonomy in the Age of AI?

- 1 Autonomous Like Us? Machinic Copies and Human Models in the Age of AI** 3
Mariafilomena Anzalone
- 2 Relational Autonomy. Rethinking AI Ethics in the Light of Feminism** 17
Stefania Achella
- 3 AI and the Relationship Between Agency, Autonomy, and Moral Patiency** 35
Paul Formosa, Inês Hipólito, and Thomas Montefiore

Part II Human and Machinic Autonomy. Ethical Implications

- 4 Automated Gatekeepers: How Recommender Systems Shape and Constrain Autonomy** 53
Hugo Cossette-Lefebvre and Natalie Stoljar
- 5 The Promises and Perils of AI: Navigating Humanity’s Most Transformative Technology** 69
Mario De Caro
- 6 Personal Autonomy and Autonomous Systems: Sameness and Difference** 81
Fiorella Battaglia
- 7 Autonomy-Enhancing Technologies? Automated Vehicles and Autonomous Mobility** 95
Fabio Fossa
- 8 Personal Avatars and Personal Autonomy** 109
Paula Sweeney

**Part III Autonomy and AI in Medicine and Health Care.
Challenges, Problems, and Opportunities**

**9 Beyond (Relational) Autonomy: AI Healthcare Through
the Lens of Care Ethics. 123**
Miranda Boldrini

**10 Balancing Autonomies: The Impact of AI on Patient and Physician
Decision-Making in Healthcare. 139**
Andrea Berber and Jelena Mijić

**11 AI in Clinical Practice: A Feyerabendian Pluralistic Model
for Physician-Patient Relationships. 155**
Deivide Garcia da Silva Oliveira

12 Big Data and the “Small X”. Patient Autonomy in the Age of AI. 171
Chantal Marazia, Vasilija Rolfes, and Fabio De Sio

**13 Humanizing Machines: Unpacking the Care Deficit in
Anthropomorphized Medical AI. 185**
Chetan Mahadev Shirvankar