

Special Issue
Engineering the Future Sociologically:
a Call to Delve into Environmental
Education Enhanced by
Technological Innovations

FUORI LUOGO

**Journal of Sociology of Territory,
Tourism, Technology**

Guest Editors

Norberto Albano
Sandro Brignone
Carmine Urciuoli



Editor in Chief: Fabio Corbisiero
Managing Editor: Carmine Urciuoli

YEAR VIII - Vol 22 - Number 1 - June 2025
FedOA - Federico II University Press
ISSN (online) 2723-9608 - ISSN (print) 2532-750X



Special Issue
Engineering the Future Sociologically:
a Call to Delve into Environmental
Education Enhanced by
Technological Innovations

FUORI LUOGO

Journal of Sociology of Territory,
Tourism, Technology

Guest editors

Norberto Albano
Sandro Brignone
Carmine Urciuoli

Editor in chief: Fabio Corbisiero
Editorial manager: Carmine Urciuoli



YEAR VIII - Vol 22 - Number 1 - June 2025
FedOA - Federico II University Press

ISSN (online) 2723-9608 – ISSN (print) 2532-750X

Summary

9. Editorial

Smart technologies and social knowledge. Where are we now?

Fabio Corbisiero

11. Engineering the Future Sociologically: Environmental Education and Advanced Technologies in the Age of Planetary Crisis

Norberto Albano, Sandro Brignone, Carmine Urciuoli

23. Mapping Technology Usage in Environmental Education

Caterina Ambrosio, Dario Chianese

35. Learning Cities and Urban Ecosystems. Digital Technologies Fostering Informal Lifelong Environmental Education in Cities and Urban Participation

Marco Ingrassia

49. ChatGPT and the Social Appropriation of AI on Discord

Vincenzo Laezza, Arianna Petrosino, Vincenzo Luise

65. AI, Big Data, and IoT for Social and Environmental Sustainability in a Digital Transformation Course

Antonio Opromolla

81. Ecoliteracy and Artificial Intelligence: Two Opposites for a Common Goal in Education

Gianfranco Rubino

3T SECTION - 3T READINGS

101. Sissa, G. (2024), *Le emissioni segrete. L'impatto ambientale dell'universo digitale*, il Mulino

Mariella Berra reads

105. Castells, M. (2024). *Digital Society*, Edward Elgar.

Giorgio Osti reads

107. Badino, M., D'Asaro F. A., Pedrazzoli, F., (2024) *Educare all'IA. La sfida didattica dell'Intelligenza Artificiale: ChatGPT e Gemini*. Sanoma.

Emanuela Ricciardi reads

INTERVIEW

113. Beyond Dichotomies: Subjectivity, Ethics, and Ontology in David J. Gunkel's Philosophy of Artificial Intelligence

Norberto Albano, Sandro Brignone, Carmine Urciuoli

FUORI LUOGO SECTION

123. Chatbots for Customer Service: the Case Study of ANAS

Giuseppina Anatriello, Massimo Carlini, Fabio Corbisiero, Maurizio Lauro, Salvatore Monaco

137. Masculinities and Caring Professions. The Case Study of Students Enrolled in University Courses in Educational Disciplines

Marianna Coppola, Giuseppe Masullo

149. Festive Expressions in Contexts of Depopulation: Carnival in the "Sicani Area" in Sicily

Alejandro Gana-Núñez

165. Italo Calvino and the Invisible Cities: Between Literature and Urban Sociology

Marxiano Melotti

183. Pandemic, culture and well-being.

A community study on the impacts of Covid-19 with respect to individual psychological well-being

Giorgio Tavano Blessi, Enzo Grossi, Matteo Colleoni

EDITOR IN CHIEF

Fabio Corbisiero (University of Naples Federico II)

✉ direttore@fuoriluogo.info

EDITORIAL MANAGER

Carmine Urciuoli

✉ caporedattore@fuoriluogo.info

SCIENTIFIC COMMITTEE

Fabio Amato (Università degli Studi di Napoli L'Orientale), Enrica Amato (Università degli Studi di Napoli Federico II), Francesco Antonelli (Università degli Studi Roma Tre), Biagio Aragona (Università degli Studi di Napoli Federico II), Arvidsson Adam Erik (Università degli Studi di Napoli Federico II), Elisabetta Bellotti (University of Manchester), Erika Bernacchi (Università degli Studi di Firenze), Kath Browne (UCD - University College Dublin), Amalia Caputo (Università degli Studi di Napoli Federico II), Letizia Carrera (Università degli Studi di Bari Aldo Moro), Gilda Catalano (Università della Calabria), Matteo Colleoni (Università degli Studi di Milano Bicocca), Linda De Feo (Università degli Studi di Napoli Federico II), Paola de Salvo (University of Perugia), Abdelhadi El Halhouli (Université Sultan Moulay Slimane – Beni Mellal – Maroc), Fiammetta Fanizza (University of Foggia), Domenica Farinella (Università degli Studi di Messina), Mariano Longo (Università del Salento), Fabiola Mancinelli (Universitat de Barcelona), Luca Marano (Università degli Studi di Napoli Federico II), Mara Maretti (Università degli Studi di Chieti Gabriele d'Annunzio), Giuseppe Masullo (Università degli Studi di Salerno), Pietro Maturi (Università degli Studi di Napoli Federico II), Antonio Maturò (Università di Bologna Alma Mater Studiorum), Claudio Milano (Universitat Autònoma de Barcelona), Khalid Mouna (Université Moulay Ismail – Mèknes - Maroc), Pierluigi Musarò (Università di Bologna Alma Mater Studiorum), Katherine O'Donnell (UCD - University College of Dublin), Giustina Orientale Caputo (Università degli Studi di Napoli Federico II), Gaia Peruzzi (Università degli Studi di Roma La Sapienza), Jason Pine (State University of New York), José Ignacio Pichardo Galán (Universidad Complutense de Madrid), Tullio Romita (Università della Calabria), Emanuele Rossi (Università degli Studi Roma Tre), Elisabetta Ruspini (Università degli Studi di Milano Bicocca), Sarah Siciliano (Università del Salento), Annamaria Vitale (Università della Calabria), Anna Maria Zaccaria (Università degli Studi di Napoli Federico II).

EDITORIAL BOARD

Amalia Caputo (Università degli Studi di Napoli Federico II)

✉ amalia.caputo@fuoriluogo.info

Rosanna Cataldo (Università degli Studi di Napoli Federico II)

✉ rosanna.cataldo@fuoriluogo.info

Linda De Feo (Università degli Studi di Napoli Federico II)

✉ linda.defeo@fuoriluogo.info

Monica Gilli (Università degli Studi di Torino)

✉ monica.gilli@fuoriluogo.info

Ilaria Marotta (Università degli Studi di Napoli Federico II)

✉ ilaria.marotta@fuoriluogo.info

Salvatore Monaco (Libera Università di Bolzano - Freie Universität Bozen)

✉ salvatore.monaco@fuoriluogo.info

Santina Musolino (Università degli Studi Roma Tre)

✉ santina.musolino@fuoriluogo.info

Francesco Santelli (Università degli Studi di Trieste)

✉ francesco.santelli@fuoriluogo.info

Redazione di Fuori Luogo

✉ redazione@fuoriluogo.info

tel. +39-081-2535883

English text editors: Pietro Maturi.

Cover by Fabio Improta. Graphic elaboration by N. Albano using Stable Diffusion IA.

EDITORE



FedOA - Federico II University Press
Centro di Ateneo per le Biblioteche "Roberto Pettorino"
Università degli Studi di Napoli Federico II

Editorial responsibility

Fedoa adopts and promotes specific guidelines on editorial responsibility, and follows COPE's Best Practice Guidelines for Journal Editors.

Authorization of the Court of Naples no. 59 of 20 December 2016.

ISSN 2723-9608 (online publication) ISSN 2532-750X (paper publication)

Articles

In evaluating the proposed works, the journal follows a peer review procedure. The articles are proposed for evaluation by two anonymous referees, once removed any element that could identify the author. Propose an article. The journal uses a submission system (open journal) to manage new proposals on the site.

<http://www.serena.unina.it/index.php/fuoriluogo>

Rights and permissions. For each contribution accepted for publication on "Fuori Luogo", the authors must return to the editorial staff a letter of authorization, completed and signed. Failure to return the letter affects the publication of the article.

The policies on the reuse of articles can be consulted on <http://www.serena.unina.it/index.php/fuoriluogo>
Fuori Luogo is one of the open access journals published under the SHARE Interuniversity Convention.

Fuori Luogo is included in the ANVUR list of Area 14 scientific journals, class A for the sociological sectors, 14/C1, 14/C2, 14/C3, 14/D1.

Fuori Luogo is indexed in: DOAJ Directory of Open Access Journals - ACNP Catalogue code n. PT03461557 - Index Copernicus International ID 67296. The journal is part of CRIS Coordinamento Riviste Italiane di Sociologia.

Fuori Luogo is included in the LOCKSS (Lots of Copies Keep Stuff Safe) network of the Public Knowledge Project (PKP PLN)

The contents are published under a Creative Commons 4.0 license.



FUORI LUOGO MEETING

Beyond Dichotomies: Subjectivity, Ethics, and Ontology in David J. Gunkel's Philosophy of Artificial Intelligence

David J. Gunkel is a globally recognized scholar in the philosophy of technology, whose research has profoundly influenced contemporary debates on the ethical and legal status of artificial intelligence and robotics. He currently serves as Presidential Research, Scholarship and Artistry Professor at Northern Illinois University and Associate Professor of Applied Ethics at Łazarski University in Warsaw. His work is distinguished by an interdisciplinary approach that bridges moral philosophy, communication theory, and emerging technologies. Author of numerous publications, Gunkel has significantly contributed to redefining the conceptual foundations of human-machine relations. Among his most recent works are: *Communicative AI: A Critical Introduction to Large Language Models* (2025, with M. Coeckelbergh), which offers a critical and accessible philosophical and theoretical overview of LLMs; *The Rights of Robots* (2024), which explores the extension of legal and moral rights to non-human entities; *Duty Now and For the Future: Communication, Ethics and Artificial Intelligence* (2023), a critical examination of the ethical responsibilities in AI development; and *ChatGPT: Deconstructing the Debate and Moving it Forward* (2023, with M. Coeckelbergh), which provides a philosophical analysis of conversational AI systems. Through these and many other contributions, Gunkel continues to set the agenda for the ethical interrogation of advanced technologies in a rapidly evolving digital society. Below is the interview conducted by the guest editors of this issue with Professor Gunkel, aimed at introducing his work to the Italian academic audience.

Interview with Prof. David J. Gunkel

Question - The increasing integration of artificial intelligence into society is not merely a matter of technological progress; rather, it compels a redefinition of the moral, epistemological, and social categories through which we comprehend agency and responsibility. The overcoming of anthropocentrism in the consideration of machines, as you have proposed, necessitates a profound reconsideration of how we conceptualize the relationships between human beings and artificial systems, as well as the criteria by which we ascribe agency and normativity to technological artefacts. Rather than proactively addressing these concerns during the design phase, debates on their implications tend to emerge only once the technology is already deeply embedded in society, making governance more challenging.

From this perspective, how can we foster a more structured and interdisciplinary dialogue between engineering and the social sciences, ensuring that AI development transcends a purely technical approach and instead becomes a deliberate and critically informed process? Specifically, how can this collaboration be designed to proactively engage with ethical and societal complexities, allowing for the anticipation and responsible governance of social and moral uncertainties, rather than merely responding to them as unforeseen externalities?

I definitely appreciate this question as it identifies what has been my principal objective as both a researcher and teacher. When we ask "How can we foster a more structured and interdisciplinary dialogue between engineering and the social sciences, ensuring that AI development transcends a purely technical approach and instead becomes a deliberate and critically informed process?" we are recognizing 1) that the opportunities and challenges of AI, robots, and other forms of seemingly intelligent and social technologies transcend disciplinary boundaries and knowledge

¹ Norberto Albano, PhD candidate, Luciano Gallino Behavioral Simulation and Educational Robotics Laboratory, Univ. of Turin; Sandro Brignone, PhD, Research Fellow, Department of Philosophy and Education Sciences, University of Turin; Carmine Urciuoli, PhD, Research Fellow, Department of Political and Social Sciences, University of Calabria

domains and 2) that responding to these opportunities and challenges will not be successful if we do not call upon and mobilize the wide range of human insight, experience, and innovation.

The unprecedented pace and scope of technological development—encompassing artificial intelligence, algorithmic governance, datafication, etc.—pose fundamental opportunities and challenges that no single discipline can adequately address in isolation. Technical advancements in machine learning and computational infrastructure, for instance, carry with them profound implications for human subjectivity, agency, ethics, and social organization. At the same time, philosophical insights into language, epistemology, and ethics are increasingly finding unexpected relevance in computational contexts. This mutual entanglement calls for a space where technologists, theorists, and practitioners can think together critically, creatively, and reflectively. Thus, it seems obvious that interdisciplinary collaboration is going to be necessary to meet the opportunities and challenges of the 21st century. But how do we do this? Or, perhaps better stated, how should this kind of productive interaction be conducted?

I think a good model can be found in what is perhaps one of the oldest methods available to us, the Platonic dialogues. Plato's dialogues routinely stage encounters between individuals representing different traditions, including philosophy, mathematics, rhetoric, politics, poetry, etc.; where no one discipline has the upper hand and insight is generated from seemingly irresolvable tensions and differences in perspective. In this way, dialogue between diverse and different domains is not simply additive but generative—capable of producing new forms of understanding that exceed what any single discipline or approach could achieve in isolation. In this respect, then, the Platonic Dialogue serves as an early and enduring paradigm of productive intellectual cooperation and pluralism. In effect, what I am suggesting is that we apply what is arguably one of the most original and ancient of philosophical methods to confront, respond to, and take responsibility for the epoch-defining challenges of our time.

But I want to add one additional item here. This procedure will only be successful if it draws on the wide range and diversity of human knowledge and experience. One of the complications with using the Platonic dialogues as a model is that this tradition is and remains Western, meaning, among other things, European and Christian. If we are to interrupt and critically address this ethnocentrism and the very real risk of intellectual colonialism, our approach to interdisciplinary dialogue also has to listen to, learn from, and engage with “non-Western” traditions (in the plural)—recognizing, of course, that this term is itself already a problem insofar as it characterizes Others negatively, thus already (and perhaps unknowingly) normalizing Western philosophy and science. An interdisciplinary dialogue that does not include other ways of seeing, thinking, and acting will remain incomplete and limited to its own ethnocentric frames of reference, modes of inquiry, and restricted range of possible solutions.

Question – *The growing pervasiveness of artificial intelligence in contemporary societies has led to a critical reconsideration of the traditional philosophical categories that distinguish subject and object, human and machine. You have argued that machines, as social robots, can no longer be interpreted merely as passive instruments but must be understood within a conceptual framework that acknowledges their potential ethical relevance and a specific form of functional agency which, rather than implying intentionality or autonomy, refers to their capacity to influence human interactions and decision-making within structured environments. From a sociological perspective, we privilege an understanding of technology not as an autonomous force but as one that is deeply embedded in and shaped by social, economic, and cultural dynamics, while still exerting a transformative influence on these very structures. Such narratives risk obscuring the socio-cultural and ethical dimensions of these transformations, reducing the debate to a fatalistic and linear perspective. How, then, can we develop a theoretical and methodological model that, while incorporating your*

critical reflections on digital subjectivity and the rethinking of the human-machine relationship, avoids both anthropocentrism and technological determinism as an independent driver of change, instead constructing a perspective that acknowledges the reciprocal and dynamic interaction between technological development and social transformation, ensuring a nuanced and contextually grounded understanding of AI's role in contemporary societies?

You ask, as I understand it, how to develop a theoretical and methodological model that avoids anthropocentrism, on the one hand, and technological determinism, on the other? For me the most potent means of intervention in this (or any other) conceptual opposition is *deconstruction*. Despite a common misperception that has (for better or worse) become something of a professional (mal)practice, what is designated by the term “deconstruction” is not synonymous with “destruction,” “disassembly,” or “dismantlement.” As Jacques Derrida, the progenitor of this method (which, strictly speaking, is not and cannot be a method) explained: “The de- of deconstruction signifies not the demolition of what is constructing itself, but rather what remains to be thought beyond the constructionist or destructionist schema” (Derrida, 1993, p. 147). Thus, “deconstruction” names a critical intervention in the system of binary oppositions by which Western philosophy and science has divided up and made sense of all that is, e.g. human/animal, person/thing, subject/object, active/passive, etc.

In deliberately and strategically intervening in these traditional conceptual oppositions, deconstruction does not choose sides in the existing debates and controversies. Instead, it institutes a double gesture of inversion—e.g. overturning an existing binary opposition—and the irruptive emergence of a new exorbitant, non-dialectical third term that cannot be contained or controlled by the existing order. Thus the way forward is not to decide between anthropocentrism or technological determinism but to identify the shared ideological and metaphysical investments that both sides already agree upon and endorse in order to come into conflict in the first place. In the process, we not only disrupt expectations of human exceptionalism and technological instrumentalism but up-end the usual expectations for who is a subject and what is an object. The goal in doing so is not to objectify the human subject or personify the technological object, but to challenge the very logic that allows us to differentiate object from subject in the first place. As Hannah Arendt (2018, p. 461) advised: “we all grow up and inherit a certain vocabulary. We then have got to examine this vocabulary.” And deconstruction names not just the examination of inherited binary terminology but a general strategy for seeing, thinking, and doing otherwise.

Question - *The debate on social robots lies at the intersection of artificial intelligence, ethics, and social theory, raising fundamental questions about their ontological status and role in human contexts. Designed to interact with us in socially recognizable and acceptable ways, these systems challenge traditional distinctions between subject and object, person and tool, with concrete ethical and legal implications. Their increasing deployment in fields such as elder care and education prompts the question of what kind of recognition - however limited - might be attributed to them, without resorting to an anthropocentric model or uncritically equating their agency with that of humans.*

Within this framework, you have argued that robots could be granted specific and limited forms of responsibility and rights, depending on their functions and social roles. Could you provide some concrete examples to clarify this position? In the case of social robots, how far should the recognition of rights and duties extend, and where should it be limited? More broadly, what criteria and principles do you consider essential for defining these forms of recognition, ensuring that social robots are neither reduced to mere tools nor mistakenly conflated with human capabilities?

Let's start with “robot rights.” Initially, when you first hear this phrase, it might strike you as not just weird but also wrong. Why would we, at this point in time, consider something like rights

for robots, when many human individuals and communities struggle for basic recognitions and protections? It's a reasonable question, but it is one that proceeds from some assumptions and even misunderstandings of what is meant by the word "rights." And if we fail to understand all the nuances here, we can be forgiven for that. In fact, even experts in law and ethics, as the American jurist Wesley Hohfeld (1920) noted over a century ago, tend to use the word "rights" in ways that are often inconsistent and not rigorously defined.

So what do we mean by "rights?" Hohfeld argues that rights are simply the set of powers, privileges, claims, or immunities that are possessed by or granted to someone or something. His point here is simple and direct: A right, like the right one has over a piece of property, can be defined and operationalized by one or more of these incidents. It can, for instance, be formulated as a claim that an owner has over and against another individual. Or it could be formulated as an exclusive privilege for use and possession that is granted to the owner. Or it could be described as a combination of these.

Hohfeld also recognized that rights are relational. A right only makes sense to the extent that it necessitates a correlative duty or obligation that is imposed on at least one other individual. This means that a right—like a claim to property ownership—means little or nothing if there is not, at the same time, some other entity who is obligated to respect this claim. Thus, rights are the way we make sense of and organize conflicting interests in social situations and circumstances.

Understood in this way, it is easy to see how and why robot rights makes sense, not (at least not principally) for the sake of the robot but for us and our shared social situations and circumstances. And it might surprise you to learn that this is not some speculation about the future. It is already happening. Robots already have rights. In November 2020, the legislature of the Commonwealth of Pennsylvania passed a bill (Senate Bill 1199) that classifies autonomous delivery robots, or what the text of the act calls "personal delivery devices" (PDD) as pedestrians in order to provide a legal framework for their deployment on city streets and sidewalks. Similar laws have been passed in a number of other jurisdictions, including the Commonwealth of Virginia, which provides the following stipulation: "a personal delivery device operating on a sidewalk or crosswalk shall have all the rights and responsibilities applicable to a pedestrian under the same circumstance." In granting this status and the rights and responsibilities that go with it to personal delivery robots, the State Legislature was not seeking to resolve or even address the big philosophical questions of robot moral standing or AI/robot personhood. It was simply seeking to scale existing law to accommodate a new kind of social entity. Robots not only can have rights, we will need them to have rights to ensure responsible integration of these technologies into our existing moral and legal systems.

Now this brings us to the last item in your question: "More broadly, what criteria and principles do you consider essential for defining these forms of recognition, ensuring that social robots are neither reduced to mere tools nor mistakenly conflated with human capabilities?" What you ask about here is the fundamental ordering principle of Western law and ethics, the division of all that is into one of two types—persons or things. This binary opposition is old, going all the way back to Roman law. As Roberto Esposito (2015, p. 1), who arguably wrote the book on this subject, explains: "If there is one assumption that seems to have organized human experience from its very beginnings it is that of a division between persons and things. No other principle is so deeply rooted in our perception and in our moral conscience..." The problem with robots and other seemingly intelligent and social artifacts is that they do not fit nicely into one or the other category. Obviously, robots are things. They are human designed and manufactured technologies that we own and possess. But these things are not quite like other things. They have social presence, they can talk and interact with us, and they successfully simulate many of the capabilities that are commonly associated with human persons. Does this therefore mean that

robots should be persons? The problem is not trying to fit robots into one category or the other. The problem is with this moral and legal ontology—this binary opposition that requires that everything be accommodated to one or the other. What we need, I believe, is a new way of thinking—a new moral and legal ontology that can scale to the novel social opportunities and challenge we now face in the face of the face plate of the robot and other forms of socially interactive technology. This is precisely the task that I have taken up and pursued in the book *Person, Thing, Robot* (MIT Press, 2023).

Question – *In your book Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond, you examine the distinction between persons and things, a principle rooted in Roman law (concerning personhood and the distinction between subjects and objects of rights), which is now being challenged not only by philosophical critiques stemming from anti-speciesism, which challenges the moral and legal privileging of human beings over other sentient entities, but also by the emergence of new technologies, such as artificial intelligence systems and robotics.*

Another critical perspective arises from the ecological crisis: recent studies indicate that the twelve wealthiest billionaires in the world, through their assets and investments, generate more carbon footprints than those produced annually by two million households. From this standpoint, the category of “person” appears to be unequally applicable even to individuals with radically different environmental impacts.

Given that some artificial intelligence systems, such as large language models, have an ecological footprint comparable to that of the economic elite rather than to that of the more impoverished segments of the population, in what ways do you believe the traditional distinction between persons and things proves inadequate in the context of contemporary sustainability crises? And how should our ethical and ontological categories be revised to account for these aspects?

The person/thing dichotomy allows us to organize everything into one of two types. The former are other moral and legal subjects who have rights and responsibilities that we need to respect and take into consideration. The latter are mere objects or what can be used and even abused as we see fit. And as Derrida (2005) points out, everything depends on the difference between “who” and “what.” The problem, especially with the climate crisis and the challenges of sustainability is that it is we—we human beings—who have invented these categories and given to ourselves the right and the privilege to divide up all that is according to this binary ordering system.

This way of thinking—which is rooted in Roman law—has gained a global foothold through colonial expansion and religious conversion. Unfortunately, this way of thinking is the philosophical scaffolding that supports human exceptionalism and justifies the exploitation of other things—and not just *things* but other human individuals and communities who were reified by Western colonial administrators; women, who until rather recently were the property—things owned by—their fathers or husbands; animals, which have been, for most of human history, little more than tools humans to use or raw material for our food, clothing, and other products; or the earth itself, which likewise has been seen as little more than a resource that can be used (and in many cases abused) for the benefit of persons with little or no consideration. Consequently, it is the person/thing binary that facilitates and supports the human exploitation of others, which is arguably one of the main driving forces behind many of the problems of the current climate crisis and sustainability. Thus, one of the philosophical opportunities with AI, robots, and other kinds of things is that they participate in the deconstruction of this binary opposition, opening up ways for being, thinking, and acting otherwise.

Question – *Contemporary models of artificial intelligence, often trained on data that reflect the stereotypes and power asymmetries embedded within the economically and technologically dominant regions of the Global North, have been described as a new form of digital colonialism. A sim-*

ilar analogy emerges in your work Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond, where you analyze the Western dichotomous thinking that opposes persons and things, instead suggesting alternative perspectives such as Ubuntu thought, which emphasizes relationality and communalism, or Confucian thought, which foregrounds relational ethics and social harmony. In your book, you discuss how robots may destabilize the traditional instrumental approach, challenging our ethical and social categories.

However, in many educational contexts, some artificial intelligence systems and robots are employed with algorithms and models that may reproduce and amplify systemic biases. How can we design and deploy these systems in ways that actively mitigate rather than reinforce existing inequalities? In what ways could these technologies be leveraged as a tool for fostering more equitable and inclusive interactions?

In order to answer this question, we first need to acknowledge how and why AI is able to generate decisions and output that can be perceived as either biased or prejudicial. This will require a short detour through the technology of AI. From the beginning—and when I say “beginning” I mean the moment the idea and term “artificial intelligence” came into being with John McCarthy et al’s plans for a summer workshop at Dartmouth College in the mid-1950’s—the field of AI divides into two different approaches: symbolic reasoning and neural networks. The former—initially theorized by Herbert Simon and Alan Newell (1958)—is based on mathematical logic and the hypothesis that intelligence consists in the manipulation of symbols. The latter is modeled on the brain, uses artificial neural networks trained on data, and machine learning methods that exploit statistical patterns and differences.

With symbolic reasoning AI or what has also been called Good Old Fashioned AI (GOF AI), the intelligence of the algorithm resides in the knowledge of the developer who translates various decision-making operations into step-by-step instructions that can be executed by a computer. If the system they design spits out erroneous or biased output, this can be identified in the lines of code and eventually corrected. Neural networks using machine learning methods are different. Here the developers simply set up the neural network and then train it on data, allowing the network to develop its own transformational operations by discovering likely patterns in the data set. In these cases, if the algorithm spits out erroneous or biased content, the problem is not in the coded instructions that set-up the neural network; the source of the problem is in the training data. And if, for example, these data, which in the case of an LLM are reams of human written content, most frequently associate the word “doctor” with the pronoun “he” and the word “nurse” with the pronoun “she,” the system will tend to produce content where doctors are gendered male and nurses female, thus perpetuating a sexist stereotype.

There are two problems here. First, because of the complexity of these systems and the sheer quantity of data on which they are trained, the developers of these systems do not know what these algorithms will do until they are in operation and start spitting out biased or prejudicial content. So the potential for bias remains largely unknown until the damage has been done. Second, efforts to correct for these problems (once they have been manifested and identified) are not as easy as they are for GOF AI. This is because the source of the bias is not in the code but “hidden” in the data. And because the quantity of data is massive, it is virtually impossible for any human being to find and remove it from the training data. Thus efforts to remediate the effects of algorithmic bias often take the form of some kind of post hoc adjustments: either a set of behavioral rules like the constitution of Claude, which filters out output that is pre-defined as unacceptable, or various forms of reinforcement learning from human feedback (RLHF) where the algorithm is put through the paces by a team of human testers and then retroactively readjusted to suppress biased or erroneous content. Both approaches have had some success, but there is no sure-fire way to ensure that the algorithm will not make mistakes, generate erroneous

or hallucinated content, or spit out biased material that users may find offensive. So, here's one way to think about the deployment of these powerful AI systems. These algorithms are like inexperienced interns. They are capable and can take over some of the heavy lifting for us, but we are going to have to check, double check, and validate everything that they produce.

Question – *In educational contexts, the introduction of robotics and artificial intelligence raises fundamental questions concerning the conceptual framework used to define robots and the ethical principles through which these technologies are interpreted. Often, the classification of robots—whether they are understood as technological instruments, autonomous agents, or even entities with quasi-personhood—is predetermined by educators and risks being conveyed to students without fostering sufficient critical examination, thereby shaping their conceptualizations in ways that may not encourage independent reflection.*

In your view, what pedagogical approach could best cultivate a more explicit and reflective awareness among students regarding the nature and theoretical status of robots? Should educators adopt a structured and theoretically grounded approach, in which they provide students with a clear and well-defined conceptual foundation, or would it be preferable to embrace a more exploratory and inclusive methodology that grants students the autonomy to develop their own perspectives?

Or do you consider this concern to be embedded within a broader epistemic and cultural framework, meaning that any conceptual or ethical perspective on robots will inevitably be shaped by the wider social and institutional context in which education takes place?

My own approach to working with students aligns with your second option. Thus I seek—in both the undergraduate and graduate courses that I teach—to cultivate “a more exploratory and inclusive methodology that grants students the autonomy to develop their own perspectives.” This approach to teaching and student learning is consistent with what philosophy is all about. The principal objective of philosophy is not to provide ready-made answers to existing questions. The task of philosophy is to interrogate and critically assess and evaluate the questions themselves. “There are,” as Slavoj Žižek (2006, p. 137) argues, “not only true or false solutions, there are also false questions. The task of philosophy is not to provide answers or solutions, but to submit to critical analysis the questions themselves, to make us see how the very way we perceive a problem is an obstacle to its solution.”

This effort at reflective self-knowledge is, it should be remembered, precisely what Immanuel Kant, the progenitor of critical philosophy, advances in the *Critique of Pure Reason*, where he deliberately avoids responding to the available questions that comprise debate in metaphysics in order to evaluate whether and to what extent the questions that are typically asked have any firm basis or foundation. Likewise, Daniel Dennett, who occupies what is often considered to be the opposite end of the philosophical spectrum from the likes of Žižek and Kant, proposes something similar. “I am a philosopher, not a scientist, and we philosophers are better at questions than answers. I haven’t begun by insulting myself and my discipline, in spite of first appearances. Finding better questions to ask, and breaking old habits and traditions of asking, is a very difficult part of the grand human project of understanding ourselves and our world” (Dennett, 1996, p. vii).

So it’s about the questions and cultivating in students not only the skill of questioning but of questioning the questions. And it is precisely for this reason, that the title of the first book in my AI/robot ethics trilogy from MIT Press is *The Machine Question* (Gunkel, 2012). The objective of this book and the two that follow it (Gunkel, 2018, 2023) is not to supply readers with definitive answers to the questions currently circulating in the AI and robot ethics literature. Instead, they have been designed to critically investigate and examine to what extent the way these existing problems and inquiries might already constitute a significant problem and difficulty. The point to all of this is to develop a mode of research

that recognizes that all questions, no matter how well formulated and carefully deployed, make exclusive decisions about what is to be included and what gets left out of consideration. The best we can do—what we have to and should do—is continually submit questions to questioning, asking not only what is given privileged status by a particular mode of inquiry and what necessarily remains excluded but also how a particular line of questioning already makes, and cannot avoid making, such decisions; what assumptions and underlying values this decision patronizes; and what consequences—ontological, epistemological, and axiological—follow from it.

References

- Arendt, H. (2018). *Thinking without a Banister: Essays in Understanding 1953–1975*. Schocken Books.
- Dennett, D. C. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
- Derrida, J. (2005). *Paper Machine*. transl. by Rachel Bowlby. Stanford University Press.
- Derrida, J. (1993). "Afterword," in *Limited Inc*. transl. by Samuel Weber. Northwestern University Press.
- Esposito, R. (2015). *Persons and Things*. Translated by Zakiya Hanafi. Cambridge: Polity.
- Gunkel, D. J. (2023). *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond*. MIT Press.
- Hohfeld, Wesley. (1920). *Fundamental Legal Conceptions as Applied in Judicial Reasoning*. Yale University Press.
- Gunkel, D. J. (2018). *Robot Rights*. MIT Press
- Gunkel, D. J. (2012). *The Machine Question: Critical Perspectives on AI, Robots and Ethics*. MIT Press.
- Simon, H.A. & Newell, A. (1958). Heuristic Problem Solving: The Next Advance in Operations Research. *Operations Research* 6(1): 1–10. <https://doi.org/10.1287/opre.6.1.1>.
- Žižek, S. (2006). Philosophy, the 'Unknown Knowns,' and the Public Use of Reason. *Topoi* 25: 137–142. <https://doi.org/10.1007/s11245-006-0021-2>.