AI, Robotics, and Humanity: Opportunities, Risks, and Implications for Ethics and Policy

Joachim von Braun, Margaret S. Archer, Gregory M. Reichberg, and Marcelo Sánchez Sorondo

Contents

Abstract

This introduction to the volume gives an overview of foundational issues in AI and robotics, looking into AI's computational basis, brain–AI comparisons, and conflicting positions on AI and consciousness. AI and robotics are changing the future of society in areas such as work, education, industry, farming, and mobility, as well as services like banking. Another important concern addressed in this volume are the impacts of AI and robotics on poor people and on inequality. These implications are being reviewed, including how to respond to challenges and how to build on the opportunities afforded by AI and robotics. An important area of new risks is robotics and AI implications for militarized conflicts. Throughout this introductory chapter and in the volume, AI/robot-human interactions, as well as the ethical and religious implications, are considered. Approaches for fruitfully managing the coexistence of humans and robots are evaluated. New forms of regulating AI and robotics are called

J. von Braun (✉)
Center for Development Research (ZEF) Bonn University, Bonn, Germany
e-mail: jvonbraun@uni-bonn.de

M. S. Archer
University of Warwick, Coventry, UK
e-mail: margaret.archer@warwick.ac.uk

G. M. Reichberg
Peace Research Institute Oslo (PRIO), Research School on Peace and Conflict | Political Science, University of Oslo, Grønland, Norway
e-mail: greg.reichberg@prio.org

M. Sánchez Sorondo
Pontifical Academy of Sciences, Vatican City, Vatican
e-mail: marcelosanchez@acdscience.va; pas@pas.va

J. von Braun et al. (eds.), *Robotics, AI, and Humanity*, https://doi.org/10.1007/978-3-030-54173-6_1

for which serve the public good but also ensure proper data protection and personal privacy.

Keywords

Artificial intelligence · Robotics · Consciousness · Labor markets · Services · Poverty · Agriculture · Militarized conflicts · Regulation

Introduction[1]

Advances in artificial intelligence (AI) and robotics are accelerating. They already significantly affect the functioning of societies and economies, and they have prompted widespread debate over the benefits and drawbacks for humanity. This fast-moving field of science and technology requires our careful attention. The emergent technologies have, for instance, implications for medicine and health care, employment, transport, manufacturing, agriculture, and armed conflict. Privacy rights and the intrusion of states into personal life is a major concern (Stanley 2019). While considerable attention has been devoted to AI/robotics applications in each of these domains, this volume aims to provide a fuller picture of their connections and the possible consequences for our shared humanity. In addition to examining the current research frontiers in AI/robotics, the contributors of this volume address the likely impacts on societal well-being, the risks for peace and sustainable development as well as the attendant ethical and religious dimensions of these technologies. Attention to ethics is called for, especially as there are also long-term scenarios in AI/robotics with consequences that may ultimately challenge the place of humans in society.

AI/robotics hold much potential to address some of our most intractable social, economic, and environmental problems, thereby helping to achieve the UN's Sustainable Development Goals (SDGs), including the reduction of climate change. However, the implications of AI/robotics for equity, for poor and marginalized people, are unclear. Of growing concern are risks of AI/robotics for peace due to their enabling new forms of warfare such as cyber-attacks or autonomous weapons, thus calling for new international security regulations. Ethical and legal aspects of AI/robotics need clarification in order to inform regulatory policies on applications and the future development of these technologies.

The volume is structured in the following four sections:

- *Foundational issues in AI and robotics*, looking into AI's computational basis, brain–AI comparisons as well as AI and consciousness.
- *AI and robotics potentially changing the future of society* in areas such as employment, education, industry, farming, mobility, and services like banking. This section also addresses the impacts of AI and robotics on poor people and inequality.
- *Robotics and AI implications for militarized conflicts* and related risks.
- *AI/robot–human interactions* and ethical and religious implications: Here approaches for managing the coexistence of humans and robots are evaluated, legal issues are addressed, and policies that can assure the regulation of AI/robotics for the good of humanity are discussed.

Foundational Issues in AI and Robotics

Overview on Perspectives

The field of AI has developed a rich variety of theoretical approaches and frameworks on the one hand, and increasingly impressive practical applications on the other. AI has the potential to bring about advances in every area of science and society. It may help us overcome some of our cognitive limitations and solve complex problems.

In health, for instance, combinations of AI/robotics with brain–computer interfaces already bring unique support to patients with sensory or motor deficits and facilitate caretaking of patients with disabilities. By providing novel tools for knowledge acquisition, AI may bring about dramatic changes in education and facilitate access to knowledge. There may also be synergies arising from robot-to-robot interaction and possible synergies of humans and robots jointly working on tasks.

While vast amounts of data present a challenge to human cognitive abilities, Big Data presents unprecedented opportunities for science and the humanities. The translational potential of Big Data is considerable, for instance in medicine, public health, education, and the management of complex systems in general (biosphere, geosphere, economy). However, the science based on Big Data as such remains empiricist and challenges us to discover the underlying causal mechanisms for generating patterns. Moreover, questions remain whether the emphasis on AI's supra-human capacities for computation and compilation mask manifold limitations

[1] The conclusions in this section partly draw on the Concluding Statement from a Conference on "Robotics, AI and Humanity, Science, Ethics and Policy", organized jointly by the Pontifical Academy of Sciences (PAS) and the Pontifical Academy of Social Sciences (PASS), 16–17 May 2019, Casina Pio IV, Vatican City. The statement is available at http://www.casinapioiv.va/content/accademia/en/events/2019/robotics/statementrobotics.html including a list of participants provided via the same link. Their contributions to the statement are acknowledged.

of current artificial systems. Moreover, there are unresolved issues of data ownership to be tackled by transparent institutional arrangements.

In the first section of this volume (Chaps. 2–5), basic concepts of AI/robotics and of cognition are addressed from different and partly conflicting perspectives. Importantly, Singer (Chap. 2) explores the difference between natural and artificial cognitive systems. Computational foundations of AI are presented by Zimmermann and Cremers (Chap. 3). Thereafter the question "could robots be conscious?" is addressed from the perspective of cognitive neuro-science of consciousness by Dehaene et al., and from a philosophical perspective by Gabriel (Chaps. 4 and 5).

Among the foundational issues of AI/robotics is the question whether machines may hypothetically attain capabilities such as consciousness. This is currently debated from the contrasting perspectives of natural science, social theory, and philosophy; as such it remains an unresolved issue, in large measure because there are many diverse definitions of "consciousness." It should not come as a surprise that the contributors of this volume are neither presenting a unanimous position on this basic issue of robot consciousness nor on a robotic form of personhood (also see Russell 2019). The concept of this volume rather is to bring the different positions together. Most contributors maintain that robots cannot be considered persons, for which reason robots will not and should not be free agents or possess rights. Some, however, argue that "command and control" conceptions may not be appropriate to human–robotic relations, and others even ask if something like "electronic citizenship" should be considered.

Christian philosophy and theology maintain that the human soul is "Imago Dei" (Sánchez Sorondo, Chap. 14). This is the metaphysical foundation according to which human persons are free and capable of ethical awareness. Although rooted in matter, human beings are also spiritual subjects whose nature transcends corporeality. In this respect, they are imperishable ("incorruptible" or "immortal" in the language of theology) and are called to a completion in God that goes beyond what the material universe can offer. Understood in this manner, neither AI nor robots can be considered persons, so robots will not and should not possess human freedom; they are unable to possess a spiritual soul and cannot be considered "images of God." They may, however, be "images of human beings" as they are created by humans to be their instruments for the good of human society. These issues are elaborated in Sect. *AI/robot–Human interactions* of the volume from religious, social science, legal, and philosophical perspectives by Sánchez Sorondo (Chap. 14), Archer (Chap. 15), and Schröder (Chap. 16).

Intelligent Agents

Zimmermann and Cremers (Chap. 3) emphasize the tremendous progress of AI in recent years and explain the conceptual foundations. They focus on the problem of induction, i.e., extracting rules from examples, which leads to the question: What set of possible models of the data generating process should a learning agent consider? To answer this question, they argue, "it is necessary to explore the notion of all possible models from a mathematical and computational point of view." Moreover, Zimmermann and Cremers (Chap. 3) are convinced that effective universal induction can play an important role in causal learning by identifying generators of observed data.

Within machine-learning research, there is a line of development that aims to identify foundational justifications for the design of cognitive agents. Such justifications would enable the derivation of theorems characterizing the possibilities and limitations of intelligent agents, as Zimmermann and Cremers elaborate (Chap. 3). Cognitive agents act within an open, partially or completely unknown environment in order to achieve goals. Key concepts for a foundational framework for AI include agents, environments, rewards, local scores, global scores, the exact model of interaction between agents and environments, and a specification of the available computational resources of agents and environments. Zimmermann and Cremers (Chap. 3) define an intelligent agent as an agent that can achieve goals in a wide range of environments.[2]

A central aspect of learning from experience is the representation and processing of uncertain knowledge. In the absence of deterministic assumptions about the world, there is no nontrivial logical conclusion that can be drawn from the past for any future event. Accordingly, it is of interest to analyze the structure of uncertainty as a question in its own right.[3] Some recent results establish a tight connection between learnability and provability, thus reducing the question of what can be effectively learned to the foundational questions of mathematics with regard to set existence axioms. Zimmermann and Cremers (Chap. 3) also point to results of "reverse mathematics," a branch of mathematical logic analyzing theorems with reference to the set of existence axioms necessary to prove them, to illustrate the implications of machine learning frameworks. They stress that artificial intelligence has advanced to a state where ethical questions and the impact on society become pressing issues, and point to the need for algorithmic transparency, accountability, and

[2] For an overview of inductive processes that are currently employed by AI-systems, see Russell (2019, pp. 285–295). The philosophical foundations of induction as employed by AI were explored inter alia by Goodman (1954).

[3] Probability-based reasoning was extended to AI by Pearl (1988).

unbiasedness. Until recently, basic mathematical science had few (if any) ethical issues on its agenda. However, given that mathematicians and software designers are central to the development of AI, it is essential that they consider the ethical implications of their work.[4] In light of the questions that are increasingly raised about the trustworthiness of autonomous systems, AI developers have a responsibility—that ideally should become a legal obligation—to create trustworthy and controllable robot systems.

Consciousness

Singer (Chap. 2) benchmarks robots against brains and points out that organisms and robots both need to possess an internal model of the restricted environment in which they act and both need to adjust their actions to the conditions of the respective environment in order to accomplish their tasks. Thus, they may appear to have similar challenges but—Singer stresses—the computational strategies to cope with these challenges are different for natural and artificial systems. He finds it premature to enter discussions as to whether artificial systems can acquire functions that we consider intentional and conscious or whether artificial agents can be considered moral agents with responsibility for their actions (Singer, Chap. 2).

Dehaene et al. (Chap. 4) take a different position from Singer and argue that the controversial question whether machines may ever be conscious must be based on considerations of how consciousness arises in the human brain. They suggest that the word "consciousness" conflates two different types of information-processing computations in the brain: first, the selection of information for global broadcasting (consciousness in the first sense), and second, the self-monitoring of those computations, leading to a subjective sense of certainty or error (consciousness in the second sense). They argue that current AI/robotics mostly implements computations similar to unconscious processing in the human brain. They however contend that a machine endowed with consciousness in the first and second sense as defined above would behave as if it were conscious. They acknowledge that such a functional definition of consciousness may leave some unsatisfied and note in closing, "Although centuries of philosophical dualism have led us to consider consciousness as unreducible to physical interactions, the empirical evidence is compatible with the possibility that consciousness arises from nothing more than specific computations." (Dehaene et al., Chap. 4, pp. . . .).

It may actually be the diverse concepts and definitions of consciousness that make the position taken by Dehaene et al. appear different from the concepts outlined by Singer (Chap. 2) and controversial to others like Gabriel (Chap. 5), Sánchez Sorondo (Chap. 14), and Schröder (Chap. 16). At the same time, the long-run expectations regarding machines' causal learning abilities and cognition as considered by Zimmermann and Cremers (Chap. 3) and the differently based position of Archer (Chap. 15) both seem compatible with the functional consciousness definitions of Dehaene et al. (Chap. 4). This does not apply to Gabriel (Chap. 5) who is inclined to answer the question "could a robot be conscious?" with a clear "no," drawing his lessons selectively from philosophy. He argues that the human being is the indispensable locus of ethical discovery. "Questions concerning what we ought to do as morally equipped agents subject to normative guidance largely depend on our synchronically and diachronically varying answers to the question of "who we are." " He argues that robots are not conscious and could not be conscious " . . . if consciousness is what I take it to be: a systemic feature of the animal-environment relationship." (Gabriel, Chap. 5, pp. . . .).

AI and Robotics Changing the Future of Society

In the second section of this volume, AI applications (and related emergent technologies) in health, manufacturing, services, and agriculture are reviewed. Major opportunities for advances in productivity are noted for the applications of AI/robotics in each of these sectors. However, a sectorial perspective on AI and robotics has limitations. It seems necessary to obtain a more comprehensive picture of the connections between the applications and a focus on public policies that facilitates overall fairness, inclusivity, and equity enhancement through AI/robotics.

The growing role of robotics in industries and consequences for employment are addressed (De Backer and DeStefano, Chap. 6). Von Braun and Baumüller (Chap. 7) explore the implications of AI/robotics for poverty and marginalization, including links to public health. Opportunities of AI/robotics for sustainable crop production and food security are reported by Torero (Chap. 8). The hopes and threats of including robotics in education are considered by Léna (Chap. 9), and the risks and opportunities of AI in financial services, wherein humans are increasingly replaced and even judged by machines, are critically reviewed by Pasquale (Chap. 10). The five chapters in this section of the volume are closely connected as they all draw on current and fast emerging applications of AI/robotics, but the balance of opportunities and risks for society differ greatly among these domains of AI/robotics applications and penetrations.

[4] The ethical impact of mathematics on technology was groundbreakingly presented by Wiener (1960).

Work

Unless channeled for public benefit, AI may raise important concerns for the economy and the stability of society. Jobs may be lost to computerized devices in manufacturing, with a resulting increase in income disparity and knowledge gaps. Advances in automation and increased supplies of artificial labor particularly in the agricultural and industrial sectors can significantly reduce employment in emerging economies. Through linkages within global value chains, workers in low-income countries may be affected by growing reliance of industries and services in higher-income countries on robotics, which could reduce the need for outsourcing routine jobs to low-wage regions. However, robot use could also increase the demand for labor by reducing the cost of production, leading to industrial expansion. Reliable estimates of jobs lost or new jobs created in industries by robots are currently lacking. This uncertainty creates fears, and it is thus not surprising that the employment and work implications of robotics are a major public policy issue (Baldwin 2019). Policies should aim at providing the necessary social security measures for affected workers while investing in the development of the necessary skills to take advantage of the new jobs created.

The state might consider to redistribute the profits that are earned from the work carried out by robots. Such redistribution could, for instance, pay for the retraining of affected individuals so that they can remain within the work force. In this context, it is important to remember that many of these new technological innovations are being achieved with support from public funding. Robots, AI, and digital capital in general can be considered as a tax base. Currently this is not the case; human labor is directly taxed through income tax of workers, but robot labor is not. In this way, robotic systems are indirectly subsidized, if companies can offset them in their accounting systems, thus reducing corporate taxation. Such distortions should be carefully analyzed and, where there is disfavoring of human workers while favoring investment in robots, this should be reversed.

Returning to economy-wide AI/robotic effects including employment, De Backer and DeStefano (Chap. 6) note that the growing investment in robotics is an important aspect of the increasing digitalization of economy. They note that while economic research has recently begun to consider the role of robotics in modern economies, the empirical analysis remains overall too limited, except for the potential employment effects of robots. So far, the empirical evidence on effects of robotics on employment is mixed, as shown in the review by De Backer and DeStefano (Chap. 6). They also stress that the effects of robots on economies go further than employment effects, as they identify increasing impacts on the organization of production in global value chains. These change the division of labor between richer and poorer economies. An important finding of De Backer and DeStefano is the negative effect that robotics may have on the offshoring of activities from developed economies, which means that robotics seem to decrease the incentives for relocating production activities and jobs toward emerging economies. As a consequence, corporations and governments in emerging economies have also identified robotics as a determinant of their future economic success. Thereby, global spreading of automation with AI/robotics can lead to faster deindustrialization in the growth and development process. Low-cost jobs in manufacturing may increasingly be conducted by robots such that fewer jobs than expected may be on offer for humans even if industries were to grow in emerging economies.

AI/Robotics: Poverty and Welfare

Attention to robot rights seems overrated in comparison to attention to implications of robotics and AI for the poorer segments of societies, according to von Braun and Baumüller (Chap. 7). Opportunities and risks of AI/robotics for sustainable development and people suffering from poverty need more attention in research and in policy (Birhane and van Dijk 2020). Especially implications for low-income countries, marginalized population groups, and women need study and consideration in programs and policies. Outcomes of AI/robotics depend upon actual designs and applications. Some examples demonstrate this crosscutting issue:

- Big Data-based algorithms drawing patterns from past occurrences can perpetuate discrimination in business practices—or can detect such discrimination and provide a basis for corrective policy actions, depending on their application and the attention given to this issue. For instance, new financial systems (fintech) can be designed to include or to exclude (Chap. 10).
- AI/robotics-aided teaching resources offer opportunities in many low-income regions, but the potential of these resources greatly depends on both the teaching content and teachers' qualifications (Léna, Chap. 9).
- As a large proportion of the poor live on small farms, particularly in Africa and South and East Asia, it matters whether or not they get access to meaningful digital technologies and AI. Examples are land ownership certification through blockchain technology, precision technologies in land and crop management, and many more (Chaps. 7 and 8).
- Direct and indirect environmental impacts of AI/robotics should receive more attention. Monitoring through smart remote sensing in terrestrial and aquatic systems can be much enhanced to assess change in biodiversity and

impacts of interventions. However, there is also the issue of pollution through electronic waste dumped by industrialized countries in low-income countries. This issue needs attention as does the carbon footprint of AI/robotics.

Effects of robotics and AI for such structural changes in economies and for jobs will not be neutral for people suffering from poverty and marginalization. Extreme poverty is on the decline worldwide, and robotics and AI are potential game changers for accelerated or decelerated poverty reduction. Information on how AI/robotics may affect the poor is scarce. Von Braun and Baumüller (Chap. 7) address this gap. They establish a framework that depicts AI/robotics impact pathways on poverty and marginality conditions, health, education, public services, work, and farming as well as on the voice and empowerment of the poor. The framework identifies points of entry of AI/robotics and is complemented by a more detailed discussion of the pathways in which changes through AI/robotics in these areas may relate positively or negatively to the livelihoods of the poor. They conclude that the context of countries and societies play an important role in determining the consequences of AI/robotics for the diverse population groups at risk of falling into poverty. Without a clear focus on the characteristics and endowments of people, innovations in AI/robotics may not only bypass them but adversely impact them directly or indirectly through markets and services of relevance to their communities. Empirical scenario building and modelling is called for to better understand the components in AI/robotics innovations and to identify how they can best support livelihoods of households and communities suffering from poverty. Von Braun and Baumüller (Chap. 7) note that outcomes much depend on policies accompanying AI and robotics. Lee points to solutions with new government initiatives that finance care and creativity (Chap. 22).

Food and Agriculture

Closely related to poverty is the influence of AI/robotics on food security and agriculture. The global poor predominantly work in agriculture, and due to their low levels of income they spend a large shares of their income on food. Torero (Chap. 8) addresses AI/robotics in the food systems and points out that agricultural production—while under climate stress—still must increase while minimizing the negative impacts on ecosystems, such as the current decline in biodiversity. An interesting example is the case of autonomous robots for farm operations. Robotics are becoming increasingly scale neutral, which could benefit small farmers via wage and price effects (Fabregas et al. 2019). AI and robotics play a growing role in all elements of food value chains, where automation is driven by labor costs as well as by demands for hygiene and food safety in processing.

Torero (Chap. 8) outlines the opportunities of new technologies for smallholder households. Small-size mechanization offers possibilities for remote areas, steep slopes or soft soil areas. Previously marginal areas could be productive again. Precision farming could be introduced to farmers that have little capital thus allowing them to adopt climate-smart practices. Farmers can be providers and consumers of data, as they link to cloud technologies using their smartphones, connecting to risk management instruments and track crop damage in real time.

Economic context may change with technologies. Buying new machinery may no longer mean getting oneself into debt thanks to better access to credit and leasing options. The reduced scale of efficient production would mean higher profitability for smallholders. Robots in the field also represent opportunities for income diversification for farmers and their family members as the need to use family labor for low productivity tasks is reduced and time can be allocated for more profit-generating activities. Additionally, robots can operate 24/7, allowing more precision on timing of harvest, especially for high-value commodities like grapes or strawberries.

Education

Besides health and caregiving, where innovations in AI/robotics have had a strong impact, in education and finance this impact is also likely to increase in the future. In education—be it in the classroom or in distance-learning systems, focused on children or on training and retraining of adults—robotics is already having an impact (Léna, Chap. 9). With the addition of AI, robotics offers to expand the reach of teaching in exciting new ways. At the same time, there are also concerns about new dependencies and unknown effects of these technologies on minds. Léna sees child education as a special case, due to it involving emotions as well as knowledge communicated between children and adults. He examines some of the modalities of teacher substitution by AI/robotic resources and discusses their ethical aspects. He emphasizes positive aspects of computer-aided education in contexts in which teachers are lacking. The technical possibilities combining artificial intelligence and teaching may be large, but the costs need consideration too. The ethical questions raised by these developments need attention, since children are extremely vulnerable human beings. As the need to develop education worldwide are so pressing, any reasonable solution which benefits from these technological advances can become helpful, especially in the area of computer-aided education.

Finance, Insurance, and Other Services

Turning to important service domains like finance and insurance, and real estate, some opportunities but also worrisome trends of applications of AI-based algorithms relying on Big Data are quickly emerging. In these domains, humans are increasingly assessed and judged by machines. Pasquale (Chap. 10) looks into the financial technology (Fintech) landscape, which ranges from automation of office procedures to new approaches of storing and transferring value, and granting credit. For instance, new services—e.g., insurance sold by the hour—are emerging, and investments on stock exchanges are conducted increasingly by AI systems, instead of by traders. These innovations in AI, other than industrial robotics, are probably already changing and reducing employment of (former) high-skill/high-income segments, but not routine tasks in manufacturing. A basis for some of the Fintech operations by established finance institutions and start-ups is the use of data sources from social media with algorithms to assess credit risk. Another area is financial institutions adopting distributed ledger technologies. Pasquale (Chap. 10) divides the Fintech landscape into two spheres, "incrementalist Fintech" and "futurist Fintech." Incrementalist Fintech uses new data, algorithms, and software to perform traditional tasks of existing financial institutions. Emerging AI/robotics do not change the underlying nature of underwriting, payment processing, or lending of the financial sector. Regulators still cover these institutions, and their adherence to rules accordingly assures that long-standing principles of financial regulation persist. Yet, futurist Fintech claims to disrupt financial markets in ways that supersede regulation or even render it obsolete. If blockchain memorializing of transactions is actually "immutable," the need for regulatory interventions to promote security or prevent modification of records may no longer be needed.

Pasquale (Chap. 10) sees large issues with futurist Fintech, which engages in detailed surveillance in order to get access to services. These can become predatory, creepy, and objectionable on diverse grounds, including that they subordinate inclusion, when they allow persons to compete for advantage in financial markets in ways that undermine their financial health, dignity, and political power (Pasquale, Chap. 10). Algorithmic accountability has become an important concern for reasons of discriminating against women for lower-paying jobs, discriminating against the aged, and stimulating consumers into buying things by sophisticated social psychology and individualized advertising based on "Phishing."[5] Pistor (2019) describes networks of obligation that even states find exceptionally difficult to break. Capital has imbricated into international legal orders that hide wealth and income from regulators and tax authorities. Cryptocurrency may become a tool for deflecting legal demands and serve the rich. Golumbia (2009) points at the potential destabilizing effects of cryptocurrencies for financial regulation and monetary policy. Pasquale (Chap. 10) stresses that both incrementalist and futurist Fintech expose the hidden costs of digital efforts to circumvent or co-opt state monetary authorities.

In some areas of innovations in AI/robotics, their future trajectories already seem quite clear. For example, robotics are fast expanding in space exploration and satellite systems observing earth,[6] in surgery and other forms of medical technology,[7] and in monitoring processes of change in the Anthropocene, for instance related to crop developments at small scales.[8] Paradigmatic for many application scenarios not just in industry but also in care and health are robotic hand-arm systems for which the challenges of precision, sensitivity, and robustness come along with safe grasping requirements. Promising applications are evolving in telemanipulation systems in a variety of areas such as healthcare, factory production, and mobility. Depending on each of these areas, sound IP standards and/or open-source innovation systems should be explored systematically, in order to shape optimal innovation pathways. This is a promising area of economic, technological, legal, and political science research.

Robotics/AI and Militarized Conflict

Robotics and AI in militarized conflicts raise new challenges for building and strengthening peace among nations and for the prevention of war and militarized conflict in general. New political and legal principles and arrangements are needed but are evolving too slowly.

Within militarized conflict, AI-based systems (including robots) can serve a variety of purposes, inter alia, extracting wounded personnel, monitoring compliance with laws of war/rules of engagement, improving situational awareness/battlefield planning, and making targeting decisions. While it is the last category that raises the most challenging moral issues, in all cases the implications of lowered barriers of warfare, escalatory dangers, as well as systemic risks must be carefully examined before AI is implemented in battlefield settings.

[5] Relevant for insights in these issues are the analyses by Akerlof and Shiller (2015) in their book on "Phishing for Phools: The Economics of Manipulation and Deception."

[6] See for instance Martin Sweeting's (2020) review of opportunities of small satellites for earth observation.

[7] For a review on AI and robotics in health see for instance Erwin Loh (2018).

[8] On assessment of fossil fuel and anthrogpogenic emissions effects on public health and climate see Jos Lelieveld et al. (2019). On new ways of crop monitoring using AI see, for instance, Burke and Lobell (2017).

Worries about falling behind in the race to develop new AI military applications must not become an excuse for short-circuiting safety research, testing, and adequate training. Because weapon design is trending away from large-scale infrastructure toward autonomous, decentralized, and miniaturized systems, the destructive effects may be magnified compared to most systems operative today (Danzig 2018). AI-based technologies should be designed so they enhance (and do not detract from) the exercise of sound moral judgment by military personnel, which need not only more but also very different types of training under the changed circumstances. Whatever military advantages might accrue from the use of AI, human agents—political and military—must continue to assume responsibility for actions carried out in wartime.

International standards are urgently needed. Ideally, these would regulate the use of AI with respect to military planning (where AI risks to encourage pre-emptive strategies), cyberattack/defense as well as the kinetic battlefields of land, air, sea, undersea, and outer space. With respect to lethal autonomous weapon systems, given the present state of technical competence (and for the foreseeable future), no systems should be deployed that function in unsupervised mode. Whatever the battlefield—cyber or kinetic—human accountability must be maintained, so that adherence to internationally recognized laws of war can be assured and violations sanctioned.

Robots are increasingly utilized on the battlefield for a variety of tasks (Swett et al., Chap. 11). Human-piloted, remote-controlled fielded systems currently predominate. These include unmanned aerial vehicles (often called "drones"), unmanned ground, surface, and underwater vehicles as well as integrated air-defense and smart weapons. The authors recognize, however, that an arms race is currently underway to operate these robotic platforms as AI-enabled weapon systems. Some of these systems are being designed to act autonomously, i.e., without the direct intervention of a human operator for making targeting decisions. Motivating this drive toward AI-based autonomous targeting systems (Lethal Autonomous Weapons, or LAWS) brings about several factors, such as increasing the speed of decision-making, expanding the volume of information necessary for complex decisions, or carrying out operations in settings where the segments of the electromagnetic spectrum needed for secure communications are contested. Significant developments are also underway within the field of human–machine interaction, where the goal is to augment the abilities of military personnel in battlefield settings, providing, for instance, enhanced situational awareness or delegating to an AI-guided machine some aspect of a joint mission. This is the concept of human–AI "teaming" that is gaining ground in military planning. On this understanding, humans and AI function as tightly coordinated parts of a multi-agent team, requiring novel modes of communication and trust. The limitations of AI must be properly understood by system designers and military personnel if AI applications are to promote more, not less, adherence to norms of armed conflict.

It has long been recognized that the battlefield is an especially challenging domain for ethical assessment. It involves the infliction of the worst sorts of harm: killing, maiming, destruction of property, and devastation of the natural environment. Decision-making in war is carried out under conditions of urgency and disorder. This Clausewitz famously termed the "fog of war." Showing how ethics are realistically applicable in such a setting has long taxed philosophers, lawyers, and military ethicists. The advent of AI has added a new layer of complexity. Hopes have been kindled for smarter targeting on the battlefield, fewer combatants, and hence less bloodshed; simultaneously, warnings have been issued on the new arms race in "killer robots," as well as the risks associated with delegating lethal decisions to increasingly complex and autonomous machines. Because LAWS are designed to make targeting decisions without the direct intervention of human agents (who are "out of the killing loop"), considerable debate has arisen on whether this mode of autonomous targeting should be deemed morally permissible. Surveying the contours of this debate, Reichberg and Syse (Chap. 12) first present a prominent ethical argument that has been advanced in favor of LAWS, namely, that AI-directed robotic combatants would have an advantage over their human counterparts, insofar as the former would operate solely on the basis of rational assessment, while the latter are often swayed by emotions that conduce to poor judgment. Several counter arguments are then presented, inter alia, (i) that emotions have a positive influence on moral judgment and are indispensable to it; (ii) that it is a violation of human dignity to be killed by a machine, as opposed to being killed by a human being; and (iii) that the honor of the military profession hinges on maintaining an equality of risk between combatants, an equality that would be removed if one side delegates its fighting to robots. The chapter concludes with a reflection on the moral challenges posed by human–AI teaming in battlefield settings, and on how virtue ethics provide a valuable framework for addressing these challenges.

Nuclear deterrence is an integral aspect of the current security architecture and the question has arisen whether adoption of AI will enhance the stability of this architecture or weaken it. The stakes are very high. Akiyama (Chap. 13) examines the specific case of nuclear deterrence, namely, the possession of nuclear weapons, not specifically for battlefield use but to dissuade others from mounting a nuclear or conventional attack. Stable deterrence depends on a complex web of risk perceptions. All sorts of distortions and errors are possible, especially in moments of crisis. AI might contribute toward reinforcing the rationality of decision-making under these conditions (easily affected by the emotional disturbances and fallacious inferences to which human beings are prone), thereby preventing an accidental launch or unintended escalation. Conversely, judgments about what does or does not fit the "national interest" are not well suited to AI (at least in its current state of development). A purely logical reasoning process based on the wrong values could have disastrous consequences, which would clearly be the case if an AI-based machine were allowed to make the launch decision (which virtually all experts would emphatically exclude), but grave problems could similarly arise if a human actor relied too heavily on AI input.

Implications for Ethics and Policies

Major research is underway in areas that define us as humans, such as language, symbol processing, one-shot learning, self-evaluation, confidence judgment, program induction, conceiving goals, and integrating existing modules into an overarching, multi-purpose intelligent architecture (Zimmermann and Cremers, Chap. 3). Computational agents trained by reinforcement learning and deep learning frameworks demonstrate outstanding performance in tasks previously thought intractable. While a thorough foundation for a general theory of computational cognitive agents is still missing, the conceptual and practical advance of AI has reached a state in which ethical and safety questions and the impact on society overall become pressing issues. For example, AI-based inferences of persons' feelings derived from face recognition data are such an issue.

AI/Robotics: Human and Social Relations

The spread of robotics profoundly modifies human and social relations in many spheres of society, in the family as well as in the workplace and in the public sphere. These modifications can take on the character of hybridization processes between the human characteristics of relationships and the artificial ones, hence between analogical and virtual reality. Therefore, it is necessary to increase scientific research on issues concerning the social effects that derive from delegating relevant aspects of social organization to AI and robots. An aim of such research should be to understand how it is possible to govern the relevant processes of change and produce those relational goods that realize a virtuous human fulfillment within a sustainable and fair societal development.

We noted above that fast progress in robotics engineering is transforming whole industries (industry 4.0). The evolution of the internet of things (IoT) with communication among machines and inter-connected machine learning results in major changes for services such as banking and finance as reviewed above. Robot–robot and human–robot interactions are increasingly intensive; yet, AI systems are hard to test and validate. This raises issues of trust in AI and robots, and issues of regulation and ownership of data, assignment of responsibilities, and transparency of algorithms are arising and require legitimate institutional arrangements.

We can distinguish between mechanical robots, designed to accomplish routine tasks in production, and AI/robotics capacities to assist in social care, medical procedures, safe and energy efficient mobility systems, educational tasks, and scientific research. While intelligent assistants may benefit adults and children alike, they also carry risks because their impact on the developing brain is unknown, and because people may lose motivation in areas where AI appears superior.

Basically robots are instruments in the perspective of Sánchez Sorondo (Chap. 14) with the term "instrument" being used in various senses. "The primary sense is clearly that of not being a cause of itself or not existing by itself." Aristotle defines being free as the one that is a cause of himself or exists on its own and for himself, i.e., one who is cause of himself (*causa sui* or *causa sui ipsius*)." From the Christian perspective, "...for a being to be free and a cause of himself, it is necessary that he/she be a person endowed with a spiritual soul, on which his or her cognitive and volitional activity is based" (Sánchez Sorondo, Chap. 14, p. 173). An artificially intelligent robotic entity does not meet this standard. As an artifact and not a natural reality, the AI/robotic entity is invented by human beings to fulfill a purpose imposed by human beings. It can become a perfect entity that performs operations in quantity and quality more precisely than a human being, but it cannot choose for itself a different purpose from what was programmed in it for by a human being. As such, the artificially intelligent robot is a means at the service of humans.

The majority of social scientists have subscribed to a similar conclusion as the above. Philosophically, as distinct from theologically, this entails some version of "human essentialism" and "species-ism" that far from all would endorse in other contexts (e.g., social constructionists). The result is to reinforce Robophobia and the supposed need to protect humankind. Margaret S. Archer (Chap. 15) seeks to put the case for potential Robophilia based upon the positive properties and powers deriving from humans and AI co-working together in synergy. Hence, Archer asks "Can Human Beings and AI Robots be Friends?" She stresses the need to foreground social change (given this is increasingly morphogenetic rather than morphostatic) for structure, culture, and agency. Because of the central role the social sciences assign to agents and their "agency" this is crucial as we humans are continually "enhanced" and have since long increased their height and longevity. Human enhancement speeded up with medical advances from ear trumpets, to spectacles, to artificial insertions in the body, transplants, and genetic modification. In short, the constitution of most adult human bodies is no longer wholly organic. In consequence, the definition of "being human" is carried further away from naturalism and human essentialism. The old bifurcation into the "wet" and the "dry" is no longer a simple binary one. If the classical distinguishing feature of humankind was held to be possession of a "soul," this was never considered to be a biological organ. Today, she argues, with the growing capacities of AI robots, the tables are turned and implicitly pose the question, "so are they not persons too?" The paradox is that the public admires the AI who defeated Chess and Go world champions. They are content with AI roles in care of the elderly, with autistic children, and in surgical interventions, none of which are purely computational feats, but the fear of artificially intelligent robots "taking over" remains and repeats Asimov's (1950) protective laws. Perceiving this as a threat alone owes much to the influence of the Arts, especially sci-fi; Robophobia dominates Robophilia in popular imagination and academia. With AI capacities now including "error-detection," "self-elaboration of their pre-programming," and "adaptation to their environment," they have the potential for *active collaboration* with humankind, in research, therapy, and care. This would entail *synergy or co-working* between humans and AI beings.

Wolfgang Schröder (Chap. 16) also addresses robot–human interaction issues, but from positions in legal philosophy and ethics. He asks what normative conditions should apply to the use of robots in human society, and ranks the controversies about the moral and legal status of robots and of humanoid robots in particular among the top debates in recent practical philosophy and legal theory. As robots become increasingly sophisticated, and engineers make them combine properties of tools with seemingly psychological capacities that were thought to be reserved for humans, such considerations become pressing. While some are inclined to view humanoid robots as more than just tools, discussions are dominated by a clear divide: What some find appealing, others deem appalling, i.e., "robot rights" and "legal personhood" for AI systems. Obviously, we need to organize human–robot interactions according to ethical and juridical principles that optimize benefit and minimize mutual harm. Schröder concludes, based on a careful consideration of legal and philosophical positions, that, even the most human-like behaving robot will not lose its ontological machine character merely by being open to "humanizing" interpretations. However, even if they do not present an anthropological challenge, they certainly present an ethical one, because both AI and ethical frameworks are artifacts *of* our societies—and therefore subject to human choice and human control, Schröder argues. The latter holds for the moral status of robots and other AI systems, too. This status remains a choice, not a necessity. Schröder suggests that there should be no context of action where a complete absence of human respect for the integrity of other beings (natural or artificial) would be morally allowed or even encouraged. Avoiding disrespectful treatment of robots is ultimately for the sake of the humans, not for the sake of the robots. Maybe this insight can contribute to inspire an "overlapping consensus" as conceptualized by John Rawls (1987) in further discussions on responsibly coordinating human-robot interactions.

Human–robot interactions and affective computing's ethical implications are elaborated by Devillers (Chap. 17). The field of social robotics is fast developing and will have wide implications especially within health care, where much progress has been made toward the development of "companion robots." Such robots provide therapeutic or monitoring assistance to patients with a range of disabilities over a long timeframe. Preliminary results show that such robots may be particularly beneficial for use with individuals who suffer from neurodegenerative pathologies. Treatment can be accorded around the clock and with a level of patience rarely found among human healthcare workers. Several elements are requisite for the effective deployment of companion robots: They must be able to detect human emotions and in turn mimic human emotional reactions as well as having an outward appearance that corresponds to human expectations about their caregiving role. Devillers' chapter presents laboratory findings on AI-systems that enable robots to recognize specific emotions and adapt their behavior accordingly. Emotional perception by humans (how language and gestures are interpreted by us to grasp the emotional states of others) is being studied as a guide to programing robots so they can simulate emotions in their interactions with humans. Some of the relevant ethical issues are examined, particularly the use of "nudges," whereby detection of a human subject's cognitive biases enables the robot to initiate, through verbal or nonverbal cues, remedial measures to affect the subject's behavior in a beneficial direction. Whether this constitutes manipulation and is open to potential abuse merits closer study.

Taking the encyclical *Laudato si'* and its call for an "integral ecology" as its starting point, Donati (Chap. 18) examines how the processes of human enhancement that have been brought about by the digital revolution (including AI and robotics) have given rise to new social relationships. A central question consists in asking how the Digital Technological Mix, a hybridization of the human and nonhuman that issues from AI and related technologies, can promote human dignity. Hybridization is defined here as entanglements and interchanges between digital machines, their ways of operating, and human elements in social practices. The issue is not whether AI or robots can assume human-like characteristics, but how they interact with humans and affect their social relationships, thereby generating a new kind of society.

Advocating for the positive coexistence of humans and AI, Lee (Chap. 22) shares Donati's vision of a system that provides for all members of society, but one that also uses the wealth generated by AI to build a society that is more compassionate, loving, and ultimately human. Lee believes it is incumbent on us to use the economic abundance of the AI age to foster the values of volunteers who devote their time and energy toward making their communities more caring. As a practical measure, they propose to explore the creation not of a universal basic income to protect against AI/robotics' labor saving and job cutting effects, but a "social investment stipend." The stipend would be given to those who invest their time and energy in those activities that promote a kind, compassionate, and creative society, i.e., care work, community service, and education. It would put the economic bounty generated by AI to work in building a better society, rather than just numbing the pain of AI-induced job losses.

Joint action in the sphere of human–human interrelations may be a model for human–robot interactions. Human–human interrelations are only possible when several prerequisites are met (Clodic and Alami, Chap. 19), inter alia: (i) that each agent has a representation within itself of its distinction from the other so that their respective tasks can be coordinated; (ii) each agent attends to the same object, is aware of that fact, and the two sets of "attentions" are causally connected; and (iii) each agent understands the other's action as intentional, namely one where means are selected in view of a goal so that each is able to make an action-to-goal prediction about the other. The authors explain how human–robot interaction must follow the same threefold pattern. In this context, two key problems emerge. First, how can a robot be programed to recognize its distinction from a human subject in the same space, to detect when a human agent is attending to something, and make judgments about the goal-directedness of the other's actions such that the appropriate predictions can be made? Second, what must humans learn about robots so they are able to interact reliably with them in view of a shared goal? This dual process (robot perception of its human counterpart and human perception of the robot) is here examined by reference to the laboratory case of a human and a robot who team up in building a stack with four blocks.

Robots are increasingly prevalent in human life and their place is expected to grow exponentially in the coming years (van Wynsberghe, Chap. 20). Whether their impact is positive or negative will depend not only on how they are used, but also and especially on how they have been designed. If ethical use is to be made of robots, an ethical perspective must be made integral to their design and production. Today this approach goes by the name "responsible robotics," the parameters of which are laid out in the present chapter. Identifying lines of responsibility among the actors involved in a robot's development and implementation, as well as establishing procedures to track these responsibilities as they impact the robot's future use, constitutes the "responsibility attribution framework" for responsible robotics. Whereas Asimov's (1950) famous "three laws of robotics" focused on the behavior of the robot, current "responsible robotics" redirects our attention to the human actors, designers, and producers, who are involved in the development chain of robots. The robotics sector has become highly complex, with a wide network of actors engaged in various phases of development and production of a multitude of applications. Understanding the different sorts of responsibility—moral, legal, backward- and forward-looking, individual and collective—that are relevant within this space, enables the articulation of an adequate attribution framework of responsibility for the robotics industry.

Regulating for Good National and International Governance

An awareness that AI-based technologies have far outpaced the existing regulatory frameworks has raised challenging questions about how to set limits on the most dangerous developments (lethal autonomous weapons or surveillance bots, for instance). Under the assumption that the robotics industry cannot be relied on to regulate itself, calls for government intervention within the regulatory space—national and international—have multiplied (Kane, Chap. 21). The author recognizes how AI technologies offer a special difficulty to any regulatory authority, given their complexity (not easily understood by nonspecialists) and their rapid pace of development (a specific application will often be obsolete by the time needed untill regulations are finally established). The various approaches to regulating AI fall into two main categories. A sectoral approach looks to identify the societal risks posed by individual technologies, so that preventive or mitigating strategies can be implemented, on the assumption that the rules applicable to AI, in say the financial industry, would be very different from those relevant to heath care providers. A cross-sectoral approach, by contrast, involves the formulation of rules (whether norms adopted by industrial consensus or laws set down by governmental authority) that, as the name implies, would have application to AI-based technologies in their generality. After surveying some domestic and international initiatives that typify the two approaches, the chapter concludes with a list of 15 recommendations to guide reflection on the promotion of societally beneficial AI.

Toward Global AI Frameworks

Over the past two decades, the field of AI/robotics has spurred a multitude of applications for novel services. A particularly fast and enthusiastic development of AI/Robotics occurred in the first and second decades of the century around industrial applications and financial services. Whether or not the current decade will see continued fast innovation and expansion of AI-based commercial and public services is an open question. An important issue is and will become even more so, how the AI innovation fields are being dominated by national strategies especially in the USA and China, or if some global arrangement for standard setting and openness can be contemplated to serve the global common good along with justifiable protection of intellectual property (IP) and fair competition in the private sector. This will require numerous rounds of negotiation concerning AI/Robotics, comparable with the development of rules on trade and foreign direct investment. The United Nations could provide the framework. The European Union would have a strong interest in engaging in such a venture, too. Civil society may play key roles from the perspective of protection of privacy.

Whether AI may serve good governance or bad governance depends, inter alia, on the corresponding regulatory environment. Risks of manipulative applications of AI for shaping public opinion and electoral interference need attention, and national and international controls are called for. The identification and prevention of illegal transactions, for instance money received from criminal activities such as drug trafficking, human trafficking or illegal transplants, may serve positively, but when AI is in the hands of oppressive governments or unethically operating companies, AI/robotics may be used for political gain, exploitation, and undermining of political freedom. The new technologies must not become instruments to enslave people or further marginalize the people suffering already from poverty.

Efforts of publicly supported development of intelligent machines should be directed to the common good. The impact on public goods and services, as well as health, education, and sustainability, must be paramount. AI may have unexpected biases or inhuman consequences including segmentation of society and racial and gender bias. These need to be addressed within different regulatory instances—both governmental and nongovernmental—before they occur. These are national and global issues and the latter need further attention from the United Nations.

The war-related risks of AI/robotics need to be addressed. States should agree on concrete steps to reduce the risk of AI-facilitated and possibly escalated wars and aim for mechanisms that heighten rather than lower the barriers of development or use of autonomous weapons, and fostering the understanding that war is to be prevented in general. With respect to lethal autonomous weapon systems, no systems should be deployed that function in an unsupervised mode. Human accountability must be maintained so that adherence to internationally recognized laws of war can be assured and violations sanctioned.

Protecting People's and Individual Human Rights and Privacy

AI/robotics offer great opportunities and entail risks; therefore, regulations should be appropriately designed by legitimate public institutions, not hampering opportunities, but also not stimulating excessive risk-taking and bias. This requires a framework in which inclusive public societal discourse is informed by scientific inquiry within different disciplines. All segments of society should participate in the needed dialogue. New forms of regulating the digital economy are called for that ensure proper data protection and personal privacy. Moreover, deontic values such as "permitted," "obligatory," and "forbidden" need to be strengthened to navigate the web and interact with robots. Human rights need to be protected from intrusive AI.

Regarding privacy, access to new knowledge, and information rights, the poor are particularly threatened because of their current lack of power and voice. AI and robotics need to be accompanied by more empowerment of the poor through information, education, and investment in skills. Policies should aim for sharing the benefits of productivity growth through a combination of profit-sharing, not by subsidizing robots but through considering (digital) capital taxation, and a reduction of working time spent on routine tasks.

Developing Corporate Standards

The private sector generates many innovations in AI/robotics. It needs to establish sound rules and standards framed by public policy. Companies, including the large corporations developing and using AI, should create ethical and safety boards, and join with nonprofit organizations that aim to establish best practices and standards for the beneficial deployment of AI/ robotics. Appropriate protocols for AI/robotics' safety need to be developed, such as duplicated checking by independent design teams. The passing of ethical and safety tests, evaluating for instance the social impact or covert racial prejudice, should become a prerequisite for the release of new AI software. External civil boards performing recurrent and transparent evaluation of all technologies, including in the military, should be considered. Scientists and engineers, as the designers of AI and robot devices, have a responsibility to ensure that their inventions and innovations are safe and can be used for moral purposes (Gibney 2020). In this context, Pope Francis has called for the elaboration of ethical guidelines for the design of algorithms, namely an "algorethics." To this he adds that "it is not enough simply to trust in the moral sense of researchers and developers of devices and algorithms. There is a need to create intermediate social bodies that can incorporate and express the ethical sensibilities of users and educators." (Pope Francis 2020). Developing and setting such standards would help in mutual learning and innovation with international spillover effects. Standards for protecting people's rights for choices and privacy also apply and may be viewed differently around the world. The general standards, however, are defined for human dignity in the UN Human Rights codex.

References

Akerlof, G. A., & Shiller, R. J. (2015). *Phishing for phools: The economics of manipulation and deception*. Princeton, NJ: Princeton University Press.

Asimov, I. (1950). Runaround. In I. Asimov (Ed.), *I, Robot*. Garden City: Doubleday.

Baldwin, R. (2019). *The globotics upheaval: Globalization, robotics, and the future of work*. New York: Oxford Umiversity Press.

Birhane, A. & van Dijk, J. (2020). *Robot rights? Let's talk about human welfare instead*. Paper accepted to the AIES 2020 conference in New York, February 2020. Doi: https://doi.org/10.1145/3375627.3375855.

Burke, M., & Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. *PNAS, 114*(9), 2189–2194; first published February 15, 2017.. https://doi.org/10.1073/pnas.1616919114.

Danzig, R. (2018). *Technology roulette: Managing loss of control as many militaries pursue technological superiority*. Washington, D.C.: Center for a New American Security. Burke M.

Fabregas, R., Kremer, M., & Schilbach, F. (2019). Realizing the potential of digital development: The case of agricultural advice. *Science, 366*, 1328. https://doi.org/10.1126/science.aay3038.

Gibney, E. (2020). The Battle to embed ethics in AI research. *Nature, 577*, 609.

Golumbia, D. (2009). *The cultural logic of computation*. Cambridge, MA: Harvard University Press.

Goodman, N. (1954). *Fact, fiction, and forecast*. London: University of London Press.

Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R. T., Haines, A., & Ramanathan, V. (2019). Effects of fossil fuel and total anthrogpogenic emission removal on public health and climate. *PNAS, 116*(15), 7192–7197. https://doi.org/10.1073/pnas.1819989116.

Loh, E. (2018). Medicine and the rise of the robots: A qualitative review of recent advances of artificial intelligence in health. *BMJ Leader, 2*, 59–63. https://doi.org/10.1136/leader-2018-000071.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco: Morgan Kaufmann.

Pistor, K. (2019). *The code of capital: How the law creates wealth and inequality*. Princeton, NJ: Princeton University Press.

Pope Francis (2020). *Discourse to the general assembly of the Pontifical Academy for Life*. Retrieved February 28, from http://press.vatican.va/content/salastampa/it/bollettino/pubblico/2020/02/28/0134/00291.html#eng.

Rawls, J. (1987). The idea of an overlapping consensus. *Oxford Journal of Legal Studies, 7*(1), 1–25.

Russell, S. (2019). *Human compatible: AI and the problem of control*. New York: Viking.

Stanley, J. (2019). *The dawn of robot surveillance*. Available via American Civil Liberties Union. Retrieved March 11, 2019, from https://www.aclu.org/sites/default/files/field_document/061119-robot_surveillance.pdf.

Sweeting, M. (2020). Small satellites for earth observation—Bringing space within reach. In J. von Braun & M. Sánchez Sorondo (Eds.), *Transformative roles of science in society: From emerging basic science toward solutions for people's wellbeing Acta Varia 25*. Vatican City: The Pontifical Academy of Sciences.

Wiener, N. (1960). Some moral and technical consequences of automation. *Science, 131*, 1355–1358. https://doi.org/10.1126/science.131.3410.1355.