

Rafael Capurro:

The Age of Artificial Intelligences: A Personal Reflection

Abstract:

The following paper presents both a historical and personal account of the societal and ethical issues arising in the development of artificial intelligence, tracking, where I was involved, the issues from the nineteen seventies onward. My own involvement in the AI narrative begins with the early discussions around whether machines can think. These first discussions, in time, evolved secondly, with the rise of the internet in the nineties, into perceptions of AI as distributed intelligence, addressing its impact on social structures including basic ethical issues arising in daily life. Thirdly, in the sweeping application of AI to all kinds of societal goals and contexts, the awareness that all natural and artificial things might be digitally connected with each other and to human agents led my further involvement in the AI narrative. Tracing this evolution from start to finish, I conclude my own narrative in the history of AI by presenting some of the future challenges for the development and use of artificial intelligences. Through the application of recent research in academia, scientific associations and political bodies, I address the possibilities for the good life, both with and without artificial intelligences.

Keywords: Artificial Intelligence, Narrative, Society, Hermeneutics, History, Ethics, Information Ethics

Outline:

1. Introduction	3
2. Can Machines Think?	3
3. A Personal Account of the History of AI	4
4. Strong and Weak AI	7
5. Distributed Intelligence	7
6. Natural and Artificial Intelligences	11
7. Conclusion: Enlightening the Digital Enlightenment	14
References	16

Author(s):

Prof. Dr. Rafael Capurro:

- Hochschule der Medien (HdM) University of Applied Science, Nobelstrasse 10, 70569 Stuttgart, Germany
- ☎ + 49 - 721 - 98 22 9 22 , ✉, rafael@capurro.de, www.capurro.de
- Relevant publications:

- Homo Digitalis. Beiträge zur Ontologie, Anthropologie und Ethik der digitalen Technik. Heidelberg: Springer VS, 2017

- Rafael Capurro, Michael Eldred and Daniel Nagel: Digital Whoness: Identity, Privacy and Freedom in the Cyberworld. Berlin: de Gruyter, 2013.

1. Introduction

What does Artificial Intelligence (AI) mean in a broad historical perspective? This is a question that has not only sociological implications but addresses the basic understanding of technology as not being purely instrumental but shaping the relationship between man and world. AI is the spirit of our time that conditions but does not determine knowledge and action. The answer to this question is a long and complex analysis going back to the roots not only of Western philosophy but also to other philosophical traditions. My aim in this paper is to recall some facts and discuss some arguments related to societal and ethical implications of AI particularly from the seventies onward where I was involved.

My own place in the narrative around AI deals first with the question that originated in cybernetics —can a machine can think— where I reference the authors such as Alan Turing, Norbert Wiener, Joseph Weizenbaum, and Hubert Dreyfus. Secondly, following the original social query presented by Norbert Wiener's Cybernetics, the rise of the Internet in the nineties evolved perceptions of AI towards issues of AI as distributed intelligence, where I participated in addressing its impact on all levels of social life. This broad societal challenge in the early day of the AI narrative was referred to as taking place in cyberspace; commonly perceived as a kind of separate sphere from the real world. This dualism soon became untenable. Thirdly, in an overarching development of digital technology for all kinds of societal goals and contexts, an awareness grew that all natural and artificial things might be digitally connected with each other as well as to human agents in what was eventually to be called the Internet of Things. Thus, in my lifetime, the understanding of the nature of AI evolved from the original Cybernetics question of whether machines can think into a question of what natural and artificial things are when they become intelligent.

2. Can Machines Think?

"I propose to consider the question, 'Can machines think?'" is the first sentence of Alan Turing's foundational paper, titled "Computing Machinery and Intelligence" (Turing, *Computing Machinery*, 433), published in 1950 in the journal *Mind*. Probably less known is Turing's report, "Intelligent Machinery" from 1948 for the National Physical Laboratory whose first sentence is similar but not identical to the later paper: "I propose to investigate the question as to whether it is possible for machinery to show intelligent behaviour." (Turing, *Intelligent Machinery*; Copeland & Proudfoot). The earlier sentence is practical and phenomenological, the later one gives rise to a theoretical epistemological question. In 1948, the mathematician Norbert Wiener wrote in his book *Cybernetics: or Control and Communication in the Animal and the Machine* the following short history of AI:

At every stage of technique since Daedalus or the Hero of Alexandria, the ability of the artificer to produce a working simulacrum of a living organism has always intrigued people. This desire to produce and to study automata has always been expressed in terms of the living technique of the age. In the days of magic, we witnessed the bizarre and sinister concept of the Golem, that figure of clay into which the Rabbi of Prague breathed life with the blasphemy of the Ineffable Name of God. In the time of Newton, the automaton becomes the clockwork music box, with their little effigies pirouetting stiffly on top. In the nineteenth century, the automaton is a glorified heat engine, burning some combustible fuel instead of the glycogen of the human muscles. Finally, the present automaton opens doors by means of photocells, or points guns to the place at which a radar beam picks up an airplane or computes the solution of a differential equation. (Wiener, *Cybernetics*, 39-40)

We can draw a history of AI via the extensive literature devoted to it (Karol Capek: *R.U.R. Rossum's Universal Robots* 1921, Stanislaw Lem: *Golem XIV*, 1981; Isaac Asimov's Three Laws of robotics became famous through the collection *I, Robot*, 1950) and as well via its representation through film (Fritz Lang: *Metropolis* 1927; Stanley Kubrick: *2001: A Space Odyssey*, 1968; Aaron Liptsadt: *Android*, 1982; Ridley Scott: *Blade Runner*, 1982; Gene Roddenberry: *Star Trek* 1987-1994, Albert Pyn: *Cyborg*, 1989; Steven Spielberg: *A.I. Artificial Intelligence*, 2001; Alex Proyas: *I, Robot*, 2004). Just as important, however, is the need to under-

stand its beginnings. The term artificial intelligence was first used in a scientific context in a workshop at Dartmouth College in 1956 at which Allen Newell, Herbert Simon, John McCarthy, Marvin Minsky and Arthur Samuel attended (McCorduck, 111-136).

The word cybernetics is of Greek origin. A *cybernetes* refers to the pilot or helmsman of a ship who faces the insecurity of sea travel, whose place it is to judge the success of a pending journey by considering a combination of factors including the weather, the sea, the robustness of the ship, and the support and loyalty of the crew. The Greeks used the term *metis* to refer to savvy intelligence, intelligence which was useful for any kind of risky endeavour. *Metis* is inclusive of skill, prudence, wisdom, cunning, and trickery (Detienne and Vernant). In a foundational text of Western thought, Aristotle writes in his *Politics* that in order to live well (*eu zen*), lifeless (*apsycha*) and living (*empsycha*) instruments (*organon*) are needed for the administration of the household (*oikia*). The rudder itself is but a lifeless instrument in regard to the piloting of a ship, while a look-out man is a living instrument. Similarly, a slave (*doulos*) is a living possession which takes the place of all other instruments. Aristotle writes:

If every instrument could accomplish its own work, following a command or anticipating it, like the statues of Daedalus, or the tripods of Hephaestus, which, says the poet, of their own accord (automatic) entered the assembly of the Gods; if, in like manner, the shuttle would weave and the plectrum touch the lyre without a hand to guide them, the master builder (architektosin) would not need servants (hypereton), nor masters (despotais) slaves (doulon). (Aristotle, *Politics*, 1253 b 25-39, revised English translation, RC)

It is somewhat ironic that Aristotle also addresses a mythical society as a utopia where work, rather than being based on the use of slaves is instead based on lifeless intelligent automata. Karl Marx quotes this text in *Das Kapital* by saying that neither Aristotle, "the greatest thinker of antiquity," nor other thinkers could comprehend

[t]he economic paradox, that the most powerful instrument for shortening labour-time, becomes the most unflinching means for placing every moment of the labourer's time and that of his family, at the disposal of the capitalist for the purpose of expanding the value of his capital. (Marx, 335; Engl. transl., 278)

The use of machines based on steam-power, electricity or digital technology creates new forms of the division of labour under, according to Marx but also to Norbert Wiener, new slave-like conditions. Wiener wrote in 1950:

Let us remember that the automatic machine, whatever we think of any feelings it may have or may not have, is the precise economic equivalent of slave labor. Any labor which competes with slave labor must accept the economic conditions of slave labor. It is completely clear that this will produce an unemployment situation, in comparison with which the present recession and even the depression of the thirties will seem a pleasant joke. (Wiener, *Cybernetics*, 162)

3. A Personal Account of the History of AI

Some of the most important and foundational developments in the IT field took place between the 1950s and the 1980s, from the first IBM computer (IBM-7101) in 1952 to the IBM Personal Computer in 1981. These developments greatly influenced my own journey. I arrived in Germany in 1971 for a two-year research appointment in a field called Documentation where I participated in theoretical courses offered by the German Documentation Society. Here I acquired practical experience at the Centre for Nuclear Documentation located on the premises of the Atomic Research Centre-Karlsruhe. This centre was the national partner of the International Nuclear Information System (INIS), created at the International Atomic Energy Agency (IAEA) as a consequence of the Sputnik crisis. Incidentally, it was the Soviets' launch of Sputnik 1 in 1957 that became the world's first artificial satellite.

The scientific and political relevance of the international exchange of information at this time was clearly stated by Alvin Weinberg, Chairman of the Panel on Science Information and Director of the Oak Ridge National Laboratory, in the 1963 report *Science, Government, and Information* (Weinberg). Weinberg's report influenced information policy in Germany from the early seventies onward and led to development of the national Information and Documentation Program (1974-1977). This program aimed at creating sixteen specialized information centres (Fachinformationszentren or FIZ), one of which was the FIZ Energy, Physics, Mathematics, later on called FIZ Karlsruhe, created in 1977 of which the former Center for Nuclear Documentation became part of.

Some years later, in 1983, STN International (Scientific and Technical Network) was created, an international online system run by FIZ Karlsruhe, Chemical Abstracts Service (CAS) and the Japan Information Center for Science and Technology (JICST). These developments created a growing awareness not only of the practical relevance of IT but also of the challenge it implied with regard to philosophical issues such as the nature of human thought and human behaviour addressed by Alan Turing. Such developments formed my original thinking on both information ethics and artificial intelligence.

It was Joseph Weizenbaum who in his book *Computer Power and Human Reason* (Weizenbaum, *Computer Power*) also raised fundamental ethical issues of computer technology. The book was published ten years after his famous *ELIZA — A Computer Program for the Study of Natural Language Communication between Man and Machine* (Weizenbaum, *ELIZA*). Weizenbaum opus magnum was a result of self-critical thinking. Herbert Simon published his *The Sciences of the Artificial* in 1969. In 1972 Hubert L. Dreyfus published the influential book *What Computers Can't Do: The Limits of Artificial Intelligence* (Dreyfus). Other studies dealing with AI followed, such as Margaret Boden's *Artificial Intelligence and Natural Man* (1977), Aaron Sloman's *The Computer Revolution in Philosophy* (1978), Daniel C. Dennett's *Brainstorms* (1978); Pamela McCorduck's *Machines Who Think* (1979); Don Ihde's *Technics and Praxis: A Philosophy of Technology* (1979); John R. Searle's *Minds, Brains, and Programs* (1980); Deborah G. Johnson's *Computer Ethics* (1985); Terry Winograd & Fernando Flores' *Understanding Computers and Cognition* (1986); P. S. Churchland's *Neurophilosophy* (1986); and Margaret Boden's (ed.) *The Philosophy of Artificial Intelligence* (1990).

A number of these scholars were deeply influenced, as I was, by the traditions of hermeneutics and phenomenology. Those who stand out include Hubert Dreyfus — I had the privilege to meet him in 1992 (Capurro, *A Long-Standing Encounter*) — and Terry Winograd and Fernando Flores. Winograd and Flores further drew my attention to AI in the eighties and early nineties after I published my post-doctoral thesis *Hermeneutics of Specialized Information* (Capurro, *Hermeneutik*). The work analysed the relationships between human understanding and interaction with computer-based information storage and retrieval.

In 1987 I was invited by German philosophers Hans Lenk and Günter Ropohl to write a contribution on the emerging field of computer ethics for a reader on *Technology and Ethics* published by Reclam. The book included contributions by Theodor W. Adorno, Hans Jonas, Kenneth D. Alpern and Alois Huning dealing particularly with ethical issues of engineering (Lenk & Ropohl). Later on, I wrote two papers on Joseph Weizenbaum whom I had the privilege to meet several times.¹

In 1987 I made a short presentation on AI at the 14th German Congress of Philosophy. The argument was, following Dreyfus, that while AI is based on explicit formalized rules, human understanding is incarnate in a body, sharing with others a common world and related to a situation that it transcends (Capurro, *Zur Kritik*). Winograd and Flores addressed this difference with regard to the design of computer systems that I analysed in a paper published in the *Informatik-Spektrum*, the German journal of Computer Science.²

¹ See Capurro, R. *Zur Computerethik; Die Verantwortbarkeit des Denkens*.

² See Capurro, *Die Informatik; Informatics*; see also: Capurro, *Die Inszenierung, Von der Künstlichen; Stellt; Ethik und Informatik*.

In 1988 I participated at the 18th World Congress of Philosophy held in Brighton where I had the opportunity to hear the speech "Men and Essences" by British Catholic, analytic and Wittgenstein expert Gertrude Elizabeth Margaret Anscombe (1919-2001) (Capurro, *Stellt die KI-Forschung*). She presented, as far as I remember, the following proof of the existence of God. There are, she said, (mathematical) essences that can be expressed in a grammar. Our capacity to produce such essences is based on our intelligence, that is to say, on our capacity to learn a language. In order to avoid a *regressus ad infinitum* we must consider that there are "intelligence or intelligences" that have created language without having received it from other beings and this is *quod omnes dicunt esse deum* (what everybody calls God). Professor Anscombe questioned my interpretation in a personal dialogue. After half an hour or so — she said several times: "please, don't interrupt me" — she asked me: "Why God? Haven't you ever heard about angels?" (Capurro, *Stellt die KI-Forschung*, 75). Having been a Jesuit for eight years (1963-1970), I was, ironically quite familiar with the concept of the angel.

It was the Jesuit theologian Karl Rahner (1904-1984) who in his book *Geist in Welt* (Spirit in the World) first opened my eyes to the relevance of angelology for dealing with the nature of human knowledge. According to Rahner, who develops his argument based on a detailed analysis of Thomas Aquinas' quaestio 89 of the *Summa Theologiae*, human knowing cannot be compared to God's knowledge, being as it is completely different to ours (Rahner).

A comparison between the archetypes of human knowledge versus the knowledge of angels allows for a useful thought experiment on the matter, where the knowledge of angels sets the *differentia specifica* of human knowledge. Being both creatures, both humans and angels have knowledge in common as *tertium comparationis*. But where humans are incarnated spirits limited to sensory experience (*conversio ad phantasmata*), condemned to return from the process of abstraction (*abstractio*) of forms, such is not the case with angels. The view of divine perennial substances (*aidiai ousiai*) separated from the finite material natural processes of becoming and decay —the Latin terminus *technicus* being *intelligentiae separatae*— goes back to Aristotle's *Metaphysics* Book Lambda (Aristotle, *Metaphysica*). Aquinas' epistemological reflection on angels and humans is based on Latin translations of Islamic philosophers such as Avicenna, who themselves translated and commented on the Greek classics. In this context of Greek-Persian-Arabic-Latin tradition, I first discovered the Greek origin (*eidos, idea, typos, morphe*) and the Latin roots of the term information (*informatio*) as well as its relationship to the concepts of messenger and message (angelos/angelia).³ Aquinas' angelology has its source in the Bible, according to which angels are immortal but not eternal creatures. In many cases they are God's messengers and not just, as in the case of Aristotle, perennial star movers⁴ (Capurro, *Ein Grinsen; Engel*).

Living in a secular and technological age, the idea of creating artificial intelligences that would even supersede human intelligence can be considered in some way in parallel to ancient and medieval thinking about divine and human intelligence, the *relata* being now human (natural) and artificial (digital). Artificial intelligences would not only enhance but eventually supersede human intelligence as the debate on transhumanism and singularity shows (Eden et al). I developed in the early nineties a critique of authors like Hans Moravec (1988) and Raymond Kurzweil (1999) in what I called cybergnosis (Capurro & Pingel), an issue that caught my attention in the early nineties as it related to the analogy between angels and computers (Capurro, *Ein Grinsen; Leben*, 78-96). According to Blaise Pascal: "L'homme n'est ni ange ni bête, et le malheur veut que qui veut faire l'ange fait la bête." (Pascal, 572) The English translation "Man is neither angel nor beast, and whoever wants to act as an angel, acts the beast" does not reflect the double meaning of the verb "faire" — to act but also to make —, although this second meaning is not the one addressed by Pascal. Many years later, in 2010, I was invited to participate at an international conference on Information Ethics: Future of Humanities, organized by the Oxford Uehiro Centre for Practical Ethics, the Uehiro Foundation on Ethics and Education, and the Carnegie Council for Ethics in International Affairs, that was held at St Cross College,

³ See Capurro, *Information; Apud Arabes*; Capurro & Holgate, *Messages and Messengers*.

⁴ See Capurro, R. Engel, *Menschen und Computer*, 2017; *Ein Grinsen ohne Katze*, 1993.

Oxford. I presented a paper with the title *Beyond Humanisms*. My argument was that Western humanisms rest on a fixation of the humanum. They are metaphysical, although they might radically differ from each other. I addressed the debate on trans- and posthumanism as follows:

The question about the humanness of the human and its "beyond" is not any more concerned with the relationship between the human and the divine as was the case with the classical humanisms in Antiquity, Renaissance and Reformation, nor with the self-introspection of the subject as in Modernity, but with the hybridization of the human, particularly through the digital medium as well as through the possibilities to change the biological substrate of the human species. A common buzzword for these issues is "human enhancement." (Capurro, *Beyond Humanisms*, 49-50)

4. Strong and Weak AI

The difference between strong and weak AI was one of the main issues discussed in the nineties.⁵ It dealt with the question of how far intelligence can be separated from its biological substrate while being (or becoming) a product of programming in the digital medium. This strong dualistic thesis became more and more problematic considering that matter matters. In other words, natural intelligence is intrinsically related to its embodiment. A bottom-up approach must consider seriously the issue of the medium. It must consider that what is crucial for the concept of artificial intelligence is not the asymptotic and unachievable approach to human intelligence but the difference created when working with another medium. When dealing with artificial intelligence(s), a key issue is to clarify the concept of artificiality.

It was the Italian sociologist Massimo Negrotti who opened my eyes to this matter.⁶ The dualism between hard- und software, underlying the strong AI thesis finds its counterpart in the metaphysical dualism between human and angelic intelligences. This dualism is portrayed by Lewis Carroll in the dialogue between Alice and the vanishing cat: "Well! I've often seen a cat without a grin," thought Alice; "but a grin without a cat! It's the most curious thing I ever saw in all my life!" (Carroll, 91; Capurro, *Ein Grinsen ohne Katze*). A few years before the technical and social revolution brought on by the advent of the Internet, the Verein Deutscher Ingenieure (VDI) (Association of German Engineers) created a group to deal with the social and ethical issues of AI, of which I was a member. A report was issued in 1992 highlighting that the classic AI model of intelligence as symbol processing and manipulation of symbolic representations implies a basic difference from that of the biological basis of human cognition. This difference should be the object of critical philosophical, psychological, linguistic and neurobiological scrutiny in order to avoid misleading analogies and to contribute to a productive use of AI in society (Cremer et al).

5. Distributed Intelligence

With the rise of the Internet in the nineties AI became an issue of distributed intelligence, that is to say, of the impact of digital technology at all levels of society giving rise to broad ethical issues that became addressed under labels as cyberethics, network ethics, information ethics and for the past ten years or so, digital ethics (Capurro, *Digital Ethics*). AI referred not so much to the creation of some kind of human-like intelligence but rather of connecting human intelligence into the digital network that has become a new human habitat analogous to what Marshall McLuhan called "the global village" in the early sixties (McLuhan).

The term cyberspace, coined by William Gibson in his 1984 novel *Neuromancer*, had now become reality with the *World Wide Web*, invented by Tim Berners-Lee in 1989. By the mid-nineties web usage was widespread thanks to the innovation of web browsers. In 1997 UNESCO organized the first international conference on

⁵ See Capurro, *Ethik und Informatik; Ein Grinsen ohne Katze*.

⁶ See Negrotti's *Theory of the Artificial; Naturoids; Reality of the Artificial*; also Capurro, *On Artificiality*.

ethical, legal and societal issues of information. My answer to the global digital challenge was the creation of the International Center for Information Ethics (ICIE) in 1999. The International Society for Ethics and IT (INSEIT) was created one year later. The first CEPE (Computer Ethics: Philosophical Inquiry) conference was held in the year 2000. The journal *Ethics and Information Technology* started in 1999. ICIE was hosted by the Center for Art and Media Karlsruhe. It started with some 20 members that grew up to some 300 hundred in the following years. ICIE organized the first international conference on intercultural issues of digital technology in 2004 sponsored by the Volkswagen Foundation. The *International Review of Information Ethics* (IRIE) was first issued in 2004, co-funded by Thomas Hausmanninger (University of Augsburg) and Felix Weil, CEO of the software company QUIBIQ (Stuttgart) who hosted the journal. ICIE and IRIE are currently overseen by the Kule Institute for Advanced Study under the leadership of Jared Bielby. I am deeply thankful to the University of Alberta and particularly to Geoffrey Rockwell, Director of the Kule Institute, and to Jared Bielby.

Social, ethical and legal issues of AI that were mainly objects of academic discussions exploded in a global context that made manifest different research agendas, cultural backgrounds and every-day lifeworlds of different societies. As the Internet took root so did concerns for an ever-growing schism between those with and those without access to its benefits. This was soon to be known as the *digital divide*. At the same time, it became evident that this was not only a technical issue. Basic questions related to privacy and democracy were at stake. The World Summit on the Information Society (WSIS) organized by the United Nations that took place 2003 in Geneva and 2005 in Tunis was a benchmark dealing at a political level with the global impact of the Internet.

Ethical issues of AI were discussed at the beginning of the new century in two EU projects in which I participated, namely ETHICBOTS (2005-2008) and ETICA (2009-2011). The ETHICBOTS (*Emerging Technoethics of Human Interaction with Communication, Bionic and Robotic Systems*) project took place under the leadership of the Italian philosopher Guglielmo Tamburrini, University "Federico II", Naples. The case-studies in the field of robotics dealt with learning robots and responsibility, military robotics (unmanned combat air vehicles), human-robots interaction (HRI) (social cognitive companions), surgery robotics and a robotic cleaning system. The case-studies in the field of bionics addressed implant technology for humans and ethics of brain computer interface technologies. The case-study in the field of AI dealt with adaptive hypermedia systems. The focus of the analysis was "the protection and promotion of human rights" (Capurro et al, ETHICBOTS, 11).

The methodological approach envisaged, first, the triaging categories of imminence, novelty, and social pervasiveness to assess the urgency of and the need for addressing techno-ethical issues, and second, a variety of ethical approaches and perspectives to represent the ethical richness of the European culture and tradition (Capurro et al, ETHICBOTS, 14). The results of the project included a paper on *Ethical Regulations on Robotics in Europe* (Nagenborg et al.) as well a book on *Ethics and Robotics* (Capurro & Nagenborg). In a contribution to the workshop L'uomo e la macchina Passato e presente (Pisa 1967-2007) organized by the Philosophy Department of the University of Pisa in May 2007, I stated the following questions to be addressed in an ethical enquiry on techno-ethics as being done in the ETHICBOTS project:

1. To whom and how should responsibility be assigned, and what should be the consequences if a human-robot integration violates or relinquishes responsibility?
2. Who is responsible for designing and maintaining an infrastructure in which information about persons is collected and processed by robots?
3. How does the possibility of invasive human-robot integration influence the concept of responsibility? This includes:
 - a. Does the fact that a human being is enhanced lead to a special kind of responsibility?
 - b. What are the consequences for those who are responsible for providing the technology used for enhancement? (Capurro, *Ethics and Robotics*, 122)

The ETICA (Ethical Issues of Emerging ICT Applications) project (2009-2011), under the leadership of the German philosopher Bernd Carsten Stahl, dealt with the following technologies: affective computing, ambient

intelligence, artificial intelligence, bioelectronics, cloud computing, future internet, human-machine symbiosis, neuroelectronics, quantum computing, robotics, virtual/augmented reality. In the *Ethical Evaluation* by Michael Nagenborg and myself we summarized the ethical issues of AI as follows:

1. *Human Dignity*: The visions of artificial persons or artificial (moral) agents with corresponding rights are to be seen as being in contrast to the emphasis given to human rights in the European Union. This might be even stronger in the case of anthropomorphic robots.
2. *Autonomy and responsibility*: The question of 'machine autonomy' does give rise to questions about human autonomy and responsibility.
3. *Privacy*: AI is one of the major building blocks of surveillance society.
4. *Cultural Diversity*: Artificial moral agents with a strong bias towards a certain cultural identity might be in contrast to a pluralistic society.
5. *Inclusion*: AI might contribute in making ICT more accessible to many people, but it might also foster the digital divide.
6. *Access to the labour market*: AI systems are likely to replace humans in certain contexts
7. *Precautionary Principle*: The precautionary principle might be invoked with regard to military applications of AI.
8. *Principle of Transparency*: The potential (bi-directional) dual use of AI systems calls for paying attention to the funding and future use of R&D in the field.
9. *Likelihood of Ethical Issues*: High. (Nagenborg & Capurro, 20-21)

In the recommendations we stated that "[t]he current research on Computer and Information Ethics is very much human-centred, which means that there is little to none research on animals or environmental issues. Therefore, we would like to encourage our colleagues to take some inspirations from the Ethics of the European Institutions and to overcome the bias towards humans" (Nagenborg & Capurro, 75).

In the *Annex to Ethical Evaluation*, Lisa Stengel and Michael Nagenborg analysed. how technology becomes an ethical issue at the level of the EU, as in the case of the work done by the *European Group on Ethics in Science and New Technologies* (EGE) for the European Commission and as well as by the National Ethics Committees. They remarked: "Since the European Community moved from a mere economic community to a political Union, it is often referred to as 'a community of values'" (Stengel & Nagenborg, 2; Capurro, Ethics and Public; Stahl).

According to *The Charter of Fundamental Rights of the European Union* such "community of values" consist of human dignity, freedom, democracy, equality, the rule of law and the respect for human rights (Nagenborg & Stengel, 1-2). The EGE issued an Opinion in 2005 on *Ethical Aspects of ICT Implants in the Human Body* (EGE, Opinion No. 20) of which Stefano Rodotà and myself were the rapporteurs. Questions were therein raised relating to AI from the perspective of the European "community of values." We summarized the issues as follows:

"We shall not lay hand upon thee". This was the promise made in the Magna Charta – to respect the body in its entirety: *Habeas Corpus*. This promise has survived technological developments. Each intervention on the body, each processing operation concerning individual data is to be regarded as related to the body as a whole, to an individual that has to be respected in its physical and mental integrity. This is a new all-round concept of individual, and its translation into the real world entails the right to full respect for a body that is nowadays both physical and electronic. In this new world, data protection fulfils the task of ensuring the "*habeas data*" required by the changed circumstances – and thereby becomes an inalienable component of civilisation, as has been the history for *habeas corpus*.

At the same time, this is a permanently unfinished body. It can be manipulated to restore functions that either were lost or were never known – only think of maiming, blindness, deafness; or, it can be stretched beyond its anthropological normality by enhancing its functions and/or adding new functions – again, for the sake of the person's welfare and/or social competitiveness, as in the case of enhanced sports skills or intelligence prostheses. We have to contend with both repairing and capaci-

ty enhancing technologies, the multiplication of body-friendly technologies that can expand and modify the concept of body care and herald the coming of 'cyborgs' – of the posthuman body. "In our societies, the body tends to become a raw material that can be modelled according to environmental circumstances". The possibilities of customised configuration undoubtedly increase, and so do the opportunities for political measures aimed at controlling the body by means of technology.

The downright reduction of our body to a device does not only enhance the trend —already pointed out— towards turning it increasingly into a tool to allow continuous surveillance of 29 individuals. Indeed, individuals are dispossessed of their own bodies and thereby of their own autonomy. The body ends up being under others' control. What can a person expect after being dispossessed of his or her own body? (EGE, Opinion No. 20, 39-30)

Stefano Rodotà (1933-2017), a famous Italian jurist and politician, published the *opus magnum Treatise of Biolaw (Trattato di Biodiritto)* edited together with Paolo Zatti, the first volume being *Field and Sources of Biolaw (Ambito e Fonti del Biodiritto)* edited by Mariachiara Tallacchini and himself. In a comprehensive contribution on "the new *habeas corpus*" Rodotà who, as far as I know, first coined the concept of "habeas data" used in the 2005 EGE Opinion, analyses key ethical and legal issues of the digitization of the human body (Rodotà 2010, 169-230). According to Rodotà, the basic right to informational self-determination overcomes the dualism between *habeas corpus*, dealing with the protection of the physical body, and *habeas data*, dealing with the protection of the electronic one. There are not different subjects to be protected but a common object, namely "the person in its different configurations, conditioned little by little in its relationship with the technologies that are not only the electronic ones." (Rodotà, 229, my translation). The EGE issued two Opinions dealing with *Ethics of Information and Communication Technologies* (EGE, Opinion No. 26) and with *Ethics of Security and Surveillance Technologies* (EGE, Opinion No. 28).

The development and use of AI devices, particularly robots, became accelerated and diversified with the impact of the Internet in all areas of social life. Turing's question from the 1950s, "can machines can think?" becomes ever more relative to the intrinsic relationship of "intelligent behaviour," Turing's formulation from 1948. Building robots and the reflection upon them becomes a social or *moral* issue, that is to say, it concerns contexts of application with specific values, customs and rules of behaviour (Latin *mores*) and a critical reflection thereupon.

In 2009 the Center for Cybernetics Research (Cybernetics) at the University of Tsukuba (Japan) organized a workshop on roboethics in which I presented a survey on roboethics (Capurro, *The Quest for Roboethics*). Makoto Nakada, one of the organizers of this workshop, presented a detailed analysis of "the meanings of 'Autonomy' of Robot Technology in the West and Japan" (Nakada, *Ethical and Critical Analysis*). Two years later we started an intercultural dialogue on roboethics. Nakada wrote:

The real intercultural ethical challenge in Japan is, I think, to ponder how robots become part of Japanese interplay between Japanese minds, which differs from the interplay in the "Far West," — particularly as it is based on the Buddhist tradition of 'self-lessness' or *Mu* — sharing a common *Ba* [place]"⁷

The nature of the self is indeed a key issue when dealing with the question of the interaction between artificial intelligences and human beings, being the case that artificial intelligences might mimic a self but, in fact, they are not (so far) a who but a what. In an interdisciplinary project organized by ACATECH (German Academy for Science and Engineering), on the question of privacy and trust on the internet, the ethics group, composed of the Australian philosopher Michael Eldred, the German lawyer Daniel Nagel and myself, developed a view of privacy and trust based on this difference. In the introduction to the ethics chapter of the final report we stated:

⁷ See Nakada & Capurro, 14; see also Nakada, Capurro, Sato; Capurro *Intercultural Roboethics*; Tzafestas, 155-175.

The concept of privacy cannot be adequately determined without its counterpart, publicness. Privacy and publicness are not properties of things, data or persons, but rather ascriptions dependent upon the specific social and cultural context. These ascriptions relate to what a person or a self (it may also be several selves) divulges about him- or herself. A self, in turn, is not a worldless, isolated subject, but a human being who is and understands herself always already connected with others in a shared world. The possibility of hiding, of displaying or showing oneself off as who one is, no matter in what way and context and to what purpose, is in this sense, as far as we know, peculiar to human beings, but precisely not as the property of a subject, but rather as a form of the interplay of a human being's life as shared with others (Capurro, Eldred, Nagel, 64).

In a special section of this chapter I analysed intercultural aspects of digitally mediated whoness, privacy and freedom in the Far East, Latin America and Africa addressing the Buddhist view of *Musi* or "denial of self" (Capurro, Intercultural Aspects, 114-116).

6. Natural and Artificial Intelligences

During the last decade of the 20th and the first decade of the 21st century breath-taking developments in the field of digital technology have taken place. *Google's* search engine was launched in 1997, *Facebook's* social network founded in 2004; and *Apple's* iPhone mobile phone released in 2007. Since then, robots have become increasingly socially relevant beyond industry applications and an entirely new conception, the Internet of Things, has in conjunction with RFID technology become a window into future possibilities. The term, Internet of Things was coined by Kevin Ashton in 1999. Ashton wrote in 2009:

We're physical, and so is our environment. Our economy, society and survival aren't based on ideas or information—they're based on things. You can't eat bits, burn them to stay warm or put them in your gas tank. Ideas and information are important, but things matter much more. [...] We need to empower computers with their own means of gathering information, so they can see, hear and smell the world for themselves, in all its random glory. RFID and sensor technology enable computers to observe, identify and understand the world—without the limitations of human-entered data. Ten years on, we've made a lot of progress, but we in the RFID community need to understand what's so important about what our technology does, and keep advocating for it. It's not just a "bar code on steroids" or a way to speed up toll roads, and we must never allow our vision to shrink to that scale. The Internet of Things has the potential to change the world, just as the Internet did. Maybe even more so. (Ashton)

The Internet, once merely an idea, is now a thing. The Internet of Things is an idea in the process of becoming a thing. It goes beyond the traditional view of artificially intelligent things called robots, envisaging the transformation of all kinds of natural and artificial things becoming digitally networked, although this might not be necessary the case. It is, I believe, the powerful IT industry with its insatiable hunger for digital data from users viewed *as* customers which supports the idea and the reality of the Internet of Things. Intelligent things might also be designed and used as stand-alone but this is bad news for IT giants. This is also true of humans as far as they might design their lives as being *always* online, what is being called *onlife*. Things and humans would eventually turn *just* into a bunch of digital data. A "who" might be confused with a "what" and a "what" might look like a "who." The difference between living online and offline does not mean a kind of Platonic dualism of separate worlds, but concerns the freedom of choice, life design and the protection of privacy. George Orwell's *Animal Farm* (Orwell) might turn into a *digital farm*. Without ideas nothing would change in the human world. What is even more relevant, I think, is that ideas make it possible for us humans to reflect on the foreseen and unforeseen changes that things bring about for non-human living and non-living, natural and artificial beings with whom we share a common world.

What are 'things' such as the Internet or the Internet of Things? They are, *prima facie*, just tools. What is a tool? In Heidegger's famous tool analysis in *Being and Time* he coined the terms "readiness-to-hand" and "presence-at-hand." When tools break down, that is to say, when they lose their readiness-to-hand, the worldly context or structure of references to which they implicitly belong, becomes manifest (Heidegger, 102ff; Capurro, Informatics and Hermeneutics). In the introduction to *Understanding Computers and Cognition* written sixty years after Heidegger's work, Winograd and Flores write how: "[...] in designing tools we are designing ways of being" (Winograd & Flores, xi). It is not due to the Internet that things are embedded in semantic and pragmatic networks but the other way around. It is because they, that is to say, we are from scratch embedded in such networks that we are able to design a digital network and interact with them and between ourselves. This means a paradigm change with regard to the way modernity conceived things as objects in the so-called "outside world" to be the correlate to an encapsulated subject. Not only humans and natural things but also artificial things become autonomous and networked agents in the digital age. But what does autonomy and action mean in each case? In a contribution to the panel on Autonomic Computing, Human Identity and Legal Subjectivity hosted by Mireille Hildebrandt and Antoinette Rouvroy at the International Conference: Computers, Privacy & Data Protection: Data Protection in a Profiled World that took place in Brussels in 2009 I wrote:

At today's early stage of these breath-taking developments, it is difficult to give a typology of the new kinds of digital and living agents and the theoretical and practical challenges arising from them. From a broad perspective, these challenges are related, on the one hand, to all kinds of robots, starting with the so-called softbots (digital robots) as well as to all kinds of physical robots – including the (still speculative) nanobots based on nanotechnology – with different degree of complexity, including all forms of imitation of human and non-human living beings (bionics). On the other hand, there are the possibilities arising from the hybridization between ICT with non-human as well as with human agents, for instance. ICT, or other technologies, can become part of living organisms, for instance as implants (EGE 2005), or vice versa. In this case, humans become (or have already become) "cyborgs" (Hayles 1999). Finally, synthetic biology allows the artificial construction of new as well as the genetic modification of living beings (EGE, Opinion No. 25; Karafyllis). (Capurro, Towards a Comparative Theory, 484)

The pervasive use of AI raises the question of the very basic understanding of technology as not being *purely* instrumental but shaping the relation between man and world. It belongs to what I call *digital ontology*, that is to say, the interpretation of the being of beings as well as of being itself from a digital perspective *as a possible one*. This ontological perspective might turn into a metaphysical world view or, politically speaking, into an ideology in the case that it becomes dogmatic, immunizing itself from critique.⁸

The Finish information security researcher Kimmo Halunen recently wrote a contribution titled "Even artificial intelligences can be vulnerable, and there are no perfect artificial intelligence applications" (Halunen). I asked him if he was the first one to use the plural noun "artificial intelligences" but he could not clarify the issue. In any case, the use of the plural noun might help to demystify the big noun AI by paying attention to a diversity of "artificial intelligence applications" making a difference with regard to other kinds of natural or artificial ones. Halunen writes:

Artificial intelligence has its own special characteristics that also make other kinds of attacks against these systems possible. Because an artificial intelligence usually attempts some kind of identification and then makes decisions based on it, the attacker may want to trick the artificial intelligence. This problem has been encountered in the fields of pattern and facial recognition in particular. Last year, it

⁸ see Capurro, Towards an Ontological Foundation of Information Ethics, 175-186; Capurro, On Floridi's Metaphysical Foundation, 167-173; Capurro, Homo Digitalis).

was published that Google's artificial intelligence algorithm was tricked into classifying a turtle as a rifle. As for facial recognition, makeup and hairstyles that fool facial recognition algorithms have been developed. Of course, people also make mistakes in identifying objects or faces, but the methods used for identification by an artificial intelligence are very different. This means that the errors made by an artificial intelligence seem bizarre to humans, because even small children can tell a turtle from a rifle, and these camouflage methods do not work against people. In an automated environment, in which artificial intelligence makes the decisions, such deceptions can be successful and may help the attacker. (Halunen)

It is energy and human needs, beliefs and desires reified in digital algorithms that move artificial intelligences (Capurro 2019). It is not primarily a question whether machines can think or how far they can be *like* human intelligence or even *better* — other machines and living beings supersede humans in many regards — but on how we might be able to live with or without them in different contexts in the life-world. Artificial intelligences or, for that matter, computer programs can break down as Winograd and Flores wrote in the eighties (Winograd & Flores 1986). The Uruguayan philosopher Fernando Flores Morador (Lund University) has done comprehensive research into what can be termed "broken technologies" (Flores Morador).

In Book 2 of his *Physics* Aristotle mentions two kinds of causes concerning what happens by chance that he calls *tyche* in the case of human action, and *automaton* in the case of natural processes (Aristotle, *Physics*, 195b 31 ff). Artificial intelligences are subject to both forms of accidental causation, due to their materiality as well as to the human intentions reified in form of algorithms that they are supposed to fulfil with more or less flexibility. They might be able to learn, anthropomorphically speaking, from their, that is to say, our failures (Wallach & Allen). Such failures can be the ones of their producers, designers and users acting intentionally or unintentionally upon them. They originate in what is supposed to take place according to foreseen or foreseeable possibilities.

In my contribution to the international conference: Artificial Intelligence & Regulation, organized by LUISS (Libera Università Internazionale degli Studi Sociali Guido Carli) held in Rome in 2018 I wrote:

Algorithms are implicitly or explicitly designed within the framework of social customs. They are embedded in cultures from scratch. According to the phenomenologist Lucas Introna, creators and users are "impressed" by algorithms (Introna 2016). The "impressionable subject," however, is not the modern subject detached from the so-called outside world, but a plurality of selves sharing a common world that is algorithmically intertwined. What is ethically at stake when dealing with algorithms becomes part of human mores? What is the nature of this entanglement between human mores and algorithms? To what extent can it be said that algorithms are, in fact, cultural? Who is responsible for the decisions taken by algorithms? To what extent is this anthropomorphic view on algorithms legitimate in order to understand what algorithms are? These are some foundational questions when dealing with the ethics of algorithms that is in an incipient state (Mittelstadt et al. 2016). [...] The present casting of ourselves as *homo digitalis* (Capurro 2017) opens the possibility of reifying ourselves algorithmically. The main ethical challenge for the inrolling digital age consists in unveiling the ethical difference, particularly when addressing the nature of algorithms and their ethical and legal regulation. (Capurro 2019)

The debate on driverless cars sometimes obfuscates basic questions on mobility that affect societies and individuals in the 21st century. At least some parts of industry seem to be interested in these issues though. I received an invitation from the *Verband der Automobilindustrie* (VDA) (German Association of the Automobile Industry) to a dialogue with the CEO of *Continental AG*, Dr. Elmar Degenhart in 2016. One year later the automobile journal *Flotten Management* invited me to a contribution on the same issue (Capurro, Fahrer entlasten). There is a growing responsibility to come into a productive dialogue towards helping academics understand the constraints and interests of the industry and industry to reflect on the questions asked by academics. This is also the case with regard to the innovations in the field of humanoid robotics as the reference volume edited by Ambarish Goswami and Prahlad Vadakkepat shows (Goswami & Vadakkepat) to which

I had the privilege of having been invited to contribute dealing with ethical issues of humanoid-human interaction (Capurro, *Ethical Issues of Humanoid*).

Last but not least, I would like to mention the previous issue of the *International Review of Information Ethics* dealing with "Ethical Issues of Networked Toys," the guest editors being Juliet Lodge and Daniel Nagel. In their introductory remarks they write:

Networked toys - Artificial guardians for little princesses or demonic plastic princes? Networked toys dominate the shelves in toy stores at a time when neither their real benefits nor their potentially latent dangers have been fully explored. Do hyper-connected toys transform the relationship between adults, the child and its environment? Do they shape their minds and predispose them to seek convenience and speedy responses rather than rely on their own autonomous capacities for critical thought?

Questions such as who really is in control arise, both of the toys —parents, third parties or even the toddlers themselves— and of data (including biometrics) that might be collected for unclear purposes and opaque destinations. For what specific or linkable purpose and above all where and to whom is data transmitted? What ethical considerations should be addressed?

Is there an actual benefit for the children themselves? Do hyper connected devices and robo-toys teach them how to handle technology or does it erode their capacity for autonomous reflection as speed and convenience are prioritised in their on-line and off-line worlds? Do such toys presage fundamental transformation of childhood and the imagined and physical worlds? (Lodge & Nagel)

7. Conclusion: Enlightening the Digital Enlightenment

The task of Information ethics in the era of artificial intelligences is to enlighten the digital enlightenment. This follows the path of thought laid out by the philosophers Max Horkheimer and Theodor W. Adorno in their influential collection of essays, *Dialectics of Enlightenment (Dialektik der Aufklärung)* (Horkheimer & Adorno). This collection addresses the ambivalence of the project(s) of enlightenment arising out of the social revolutions of the nineteenth century, reflecting on the dialectics between mythology and science that characterized European Enlightenment particularly in the eighteenth century. Enlightenment must take care of this ambivalence that might revert digital enlightenment into digital mythology. This narrative highlights how the changing meanings of the concept of artificial intelligence(s) over the last century depend both on the state-of-the-art of digital technology as well as the different contexts in which it has been used. Looking back on my personal experience in the AI narrative since the early seventies and the changing academic debates in the years that followed, I dare not forecast beyond the most imminent challenges of the near future.

The task of taming the digital chaos through different kinds of national and international regulations is still very much in the early stages and is dependent on how the awareness of these issues take root across the globe. Enlightened awareness addresses several problems such as ecological issues, sustainability, taxation, state regulation, fake news, cyber wars, digital capitalism, digital colonialism, social justice, surveillance society, digital addiction, the future of work, and who we are as cybercitizens in the digital age. Toni Samek and Lynette Schultz organized an Information Ethics Roundtable at the University of Alberta in 2014 dealing with this last issue (Samek & Schultz; Capurro, *Citizenship*).

To lead the enlightenment of the digital enlightenment we look to the universities, research institutions, scientific associations, governments and the media. As a paramount example of a scientific association leading in the field of enlightening the digital enlightenment, I would like to mention the Institution of Electrical and Electronics Engineers Standards Association (IEEE), which initiated the Global Initiative of Autonomous and Intelligent Systems under the leadership of managing director Konstantinos Karachalios and John Havens, who led as Executive Director of the Initiative. The Chair of the International Center for Information Ethics, Jared Bielby, managed the Committee for Classical Ethics for the initiative, assisted by Rachel Fischer,

a leading practitioner in the field of information ethics in South Africa. The final report represents the collective input of several hundred participants from six continents who are thought to be leaders from academia, industry, civil society, policy and government (IEEE). I quote the introduction *in extenso*:

The task of the Committee for Classical Ethics in Autonomous and Intelligent Systems is to apply classical ethics methodologies to considerations of algorithmic design in autonomous and intelligent systems (A/IS) where machine learning may or may not reflect ethical outcomes that mimic human decision-making. To meet this goal, the Committee has drawn from classical ethics theories as well as from the disciplines of machine ethics, information ethics, and technology ethics. As direct human control over tools becomes, on one hand, further removed, but on the other hand, more influential than ever through the precise and deliberate design of algorithms in self-sustained digital systems, creators of autonomous systems must ask themselves how cultural and ethical presumptions bias artificially intelligent creations, and how these created systems will respond based on such design. By drawing from over two thousand years' worth of classical ethics traditions, the Classical Ethics in Autonomous and Intelligent Systems Committee will explore established ethics systems, addressing both scientific and religious approaches, including secular philosophical traditions such as utilitarianism, virtue ethics, deontological ethics and religious and culture-based ethical systems arising from Buddhism, Confucianism, African Ubuntu traditions, and Japanese Shinto influences toward an address of human morality in the digital age. In doing so the Committee will critique assumptions around concepts such as good and evil, right and wrong, virtue and vice and attempt to carry these inquiries into artificial systems decision-making processes. Through reviewing the philosophical foundations that define autonomy and ontology, the Committee will address the potential for autonomous capacity of artificially intelligent systems, posing questions of morality in amoral systems, and asking whether decisions made by amoral systems can have moral consequences. Ultimately, it will address notions of responsibility and accountability for the decisions made by autonomous systems and other artificially intelligent technologies. (IEEE)

At the political level I highlight and support the recent activities of the European Union, particularly the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe (European Union). The age of artificial intelligences has just started.

References

- Aristotle. *Metaphysica*. Ed. W. Jaeger. London: Oxford University Press, 1973.
- Aristotle. *Physica*. Ed. W. D. Ross. Oxford: The Clarendon Press, 1950.
- Aristotle. *The Politics of Aristotle*. Ed. W.L. Newman. Oxford: The Clarendon Press, 1950.
- Ashton, K. The 'Internet of Things' Thing. In the real world, things matter more than ideas. In: *RFID Journal*, 2009.
- Boden, M. A. *Artificial Intelligence and Natural Man*. Pepperell, MA: Branch Line Press, 1977.
- Boden, M. A. (ed.) *The Philosophy of Artificial Intelligence*. New York: Oxford University Press, 1990.
- Capurro, R. *Apud Arabes. Notes on Greek, Latin, Arabic, Persian, and Hebrew Roots of the Concept of Information*, 2018. <http://www.capurro.de/iran.html>
- Capurro, R. *Beyond Humanisms*. In: T. Nishigaki, T. Takenouchi (eds.): *Information Ethics. The Future of the Humanities*, Nagoya City: V2 Solution Publisher, 2012, pp. 26-74.
- Capurro, R. *Citizenship in the Digital Age*. In: T. Samek, L. Schultz (eds.): *Information Ethics, Globalization and Citizenship. Essays on Ideas to Praxis*. Jefferson NC: McFarland, 2017, pp. 11-30.
- Capurro, R. *Der Kongress. Rafael Capurro über den XVIII. Weltkongress für Philosophie in Brighton, GB*. In: *Information Philosophie*, 2, pp. 74-82.
- Capurro, R. *Die Informatik und das hermeneutische Forschungsprogramm*. In: *Informatik-Spektrum* 10, 1987, pp. 329-333.
- Capurro, R. *Die Inszenierung des Denkens*. In: *Mensch – Natur – Gesellschaft* Jg. 5, Heft 1, 1988, pp. 18-26.
- Capurro, R. *Die Verantwortbarkeit des Denkens. Künstliche Intelligenz aus ethischer Sicht*. In: *Forum für interdisziplinäre Forschung* 1, 1998, pp. 15-21.
- Capurro, R. *Digital Ethics*. In: *The Academy of Korean Studies (ed.): 2009 Civilization and Peace*, Korea: The Academy of Korean Studies, 2010, pp. 203-214.
- Capurro, R. *Ein Grinsen ohne Katze. Von der Vergleichbarkeit zwischen 'künstlicher Intelligenz' und 'getrennten Intelligenzen'*. In: *Zeitschrift für philosophische Forschung*, 47, 1993, pp. 93-102.
- Capurro, R. *Enculturating Algorithms*. In: *Nanoethics*, 2019, pp. 1-7. <https://doi.org/10.1007/s11569-019-00340-9> (pdf)
- Capurro, R. *Engel, Menschen und Computer. Zur Relevanz der thomistischen Engellehre für die philosophischen Anthropologie*, 2017. <http://www.capurro.de/engel.html>
- Capurro, R. *Ethical Issues of Humanoid-Human Interaction*. In: A. Goswami, P. Vadakkepat (eds.): *Humanoid Robotics: A Reference*. Heidelberg: Springer, 2019, pp. 2421-2436. https://link.springer.com/referenceworkentry/10.1007%2F978-94-007-6046-2_127
- Capurro, R. *Ethics and Public Policy in Europe*. In: S. Rodotà, M. Tallachini (eds.): *Ambito e fonti del biodiritto*. Milano: Giuffrè, 2010, pp. 849-860.

Capurro, R. *Ethics and Robotics*. In: R. Capurro, M. Nagenborg (eds.). *Ethics and Robotics*. Heidelberg: Akademische Verlagsgesellschaft, 2009, pp.117-123.

Capurro, R. *Ethik und Informatik*. In: *Informatik-Spektrum* 13, 1990, pp. 311-320.

Capurro, R. *Fahrer entlasten nicht ersetzen*. In: *Flotten management* 1, 2017, 76-77 <https://www.flotte.de/magazine/flottenmanagement-magazin/2017/1/71/innovation--technik/5759/fahrer-entlasten-nicht-ersetzen.html>

Capurro, R. *Homo Digitalis. Beiträge zur Ontologie, Anthropologie und Ethik der digitalen Technik*. Heidelberg: Springer, 2017.

Capurro, R. *Hermeneutik der Fachinformation*. Freiburg/München: Alber, 1986.

Capurro, R. *Informatics and Hermeneutics*. In: Ch. Floyd, H. Züllighoven, R. Budde, R. Keil-Slawik (eds.): *Software Development and Reality Construction*. Berlin: Springer, 1992, pp. 363-375.

Capurro, R. *Information. Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs*. München: Saur, 1978.

Capurro, R. *Intercultural Aspects of Digitally Mediated Whoness, Privacy and Freedom*. In: R. Capurro, R., M. Eldred, D. Nagel: 'IT and Privacy from an Ethical Perspective: Digital Whoness: Identity, Privacy and Freedom in the Cyberworld'. In: Johannes Buchmann (ed.) *Internet Privacy – Eine multidisziplinäre Bestandsaufnahme. A Multidisciplinary Analysis*. Acatech Studie, Berlin, 2012, pp. 113-122.

Capurro, R. *Intercultural Roboethics for a Robot Age*. In: M. Nakada, R. Capurro, K. Sato (eds.) (2017). *Critical Review of Information Ethics and Roboethics in East and West. Master's and Doctoral Program in International and Advanced Japanese Studies, Research Group for "Ethics and Technology in the Information Era"*, University of Tsukuba, ISSN 2432-5414, 13-18.

Capurro, R. *Leben im Informationszeitalter*. Berlin: Akademie Verlag, 1995.

Capurro, R. *A Long-Standing Encounter*. In: *AI & Society*, 2018. pp. 1-2.

Capurro, R. *On Artificiality*. In: IMES (Istituto Metodologico Economico Statistico) Laboratory for the Culture of the Artificial, Università di Urbino, IMES-LCA WP, 1995.

Capurro, R. *On Floridi's Metaphysical Foundation of Information Ecology*. In: *Ethics and Information Technology*, 10 (2-3), 2008, pp. 167-173.

Capurro, R. *The Quest for Roboethics: A Survey*. In: T. Kimura, M. Nakada, K. Suzuki, Y. Sankai (eds.): *Cybernetics Technical Reports. Special Issue on Roboethics*. University of Tsukuba, CYB-2011-001 - CYB-2011-008, 2011, pp. 39-59.

Capurro, R. *Stellt die KI-Forschung den Kernbereich der Informatik dar?* In: J. Retti & K. Leidlmaier (eds.): *5. Österreichische Artificial-Intelligence-Tagung*. Berlin: Springer 1989, S. 415-421.

Capurro, R. *Towards a Comparative Theory of Agents*. In: *AI & Society*, 27 (4), 2012, pp. 479-488.

Capurro, R. *Towards an Ontological Foundation of Information Ethics*. In: *Ethics and Information Technology* 8 (4), 2006, pp. 175-186

Capurro, R. *Von der Künstlichen Intelligenz als einem ästhetischen Phänomen. Eine kritische Reflexion in Kantischer Absicht*, 1988.

Capurro, R. *Zur Computerethik. Ethische Fragen der Informationsgesellschaft.* In: H. Lenk, G. Ropohl (eds.): *Technik und Ethik.* Stuttgart: Reclam, 1987, pp. 259-273,

Capurro, R. *Zur Kritik der künstlichen Vernunft. Oder über die Einheit und Vielheit des Intellekts.* In: O. Marquard, P. Probs (eds.): *Proceedings of the 14th German Congress of Philosophy, Giessen, September 21-16, 1987.*

Capurro, R., Eldred, M., Nagel, D. *'IT and Privacy from an Ethical Perspective: Digital Whoness: Identity, Privacy and Freedom in the Cyberworld.* In: J. Buchmann (ed.) *Internet Privacy – Eine multidisziplinäre Bestandsaufnahme. A Multidisciplinary Analysis.* Acatech Studie, Berlin, 2012, pp. 63-142. <https://www.acatech.de/Publikation/internet-privacy-eine-multidisziplinare-bestandsaufnahme/>

Capurro, R., Holgate, J. (eds.) *Messages and Messengers. Angeletics as an Approach to the Phenomenology of Communication.* München: Fink, 2011.

Capurro, R., Nagenborg, M. (eds.) *Ethics and Robotics.* Heidelberg: Akademische Verlagsgesellschaft, 2009.

Capurro, R., Pingel, Ch. *Ethical Issues of Online Communication Research.* In: *Ethics and Information Technology* 4 (3), 2002, pp. 189-194.

Capurro, R., Tamburrini, G., Weber J. *ETHICBOTS, D 5: Techno-ethical Case-Studies in Robotics, Bionics, and Related AI Agent Information Technologies,* 2008. https://www.researchgate.net/publication/258833649_Techno-Ethical_Case-Studies_in_Robotics_Bionics_and_Related_AI_Agent_Technologies

Carroll, L. *The Annotated Alice.* New York: Bramhall House, 1965.

Churchland, P.S. *Neurophilosophy: Toward a unified science of the mind-brain.* Cambridge, MA: MIT Press, 1986.

Dennett, D. C. *Brainstorms: Philosophical Essays on Mind and Psychology.* Denver Colorado: Bradford Publishing, 1978.

Detienne, M., Vernant, J.-P. *Les ruses de l'intelligence. La mêtis des Grecs. (Cunning Intelligence in Greek Culture and Society).* Paris: Flammarion, 1974.

Dreyfus, H.L. *What Computers Can't Do. The Limits of Artificial Intelligence.* New York: MIT Press, 1972.

Eden, A.H., Moor, J.H., Soraker, J.H., Steinhart, E. (eds.) *Singularity Hypotheses. A Scientific and Philosophical Assessment.* Heidelberg: Springer, 2012.

European Group on Ethics in Science and New Technologies (EGE) *Opinion No. 20: Ethical Aspects of ICT Implants in the Human Body,* 2005. <http://ec.europa.eu/research/ege/index.cfm?pg=reports>

European Group on Ethics in Science and New Technologies (EGE) *Opinion No. 25: Ethics of Synthetic Biology,* 2009. <http://ec.europa.eu/research/ege/index.cfm?pg=reports>

European Group on Ethics in Science and New Technologies (EGE) *Opinion No. 26: Ethics of information and communication technologies,* 2012. <http://ec.europa.eu/research/ege/index.cfm?pg=reports>

European Group on Ethics in Science and New Technologies (EGE) *Opinion No. 28: Ethics of Security and Surveillance Technologies,* 2014. <http://ec.europa.eu/research/ege/index.cfm?pg=reports>

- European Union. *Communication Artificial Intelligence for Europe*. Brussels, 2018. <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>
- Flores Morador, F. *Encyclopedia of Broken Technologies - The Humanist as Engineer (October 23, 2015)*. *Broken Technologies. The Humanist as Engineer*. Ver 3.0, Lund University, Revised 2015. Available at SSRN: <https://ssrn.com/abstract=2679102>
- Goswami, A., Vadakkepat, P. (eds.) *Humanoid Robotics: A Reference*. Heidelberg: Springer, 2019. <https://link.springer.com/referencework/10.1007/978-94-007-6046-2>
- Halunen, K. *Even artificial intelligences can be vulnerable, and there are no perfect artificial intelligence applications*. VTT Technical Research Centre of Finland, 23/08/2018. <https://vttblog.com/2018/08/23/even-artificial-intelligences-can-be-vulnerable-and-there-are-no-perfect-artificial-intelligence-applications/>
- Heidegger, M. *Being and Time*. Oxford: Blackwell. Transl. J. Macquarrie, E. Robinson, 1987.
- Horkheimer, M., Adorno, Th. W. *Dialektik der Aufklärung*. Frankfurt am Main: Fischer, 1975.
- IEEE Standards Association. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, 2016. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html
- Ihde, Don. *Technics and Praxis: A Philosophy of Technology*. Dordrecht: D. Reidel, 1979.
- Johnson, D.G. *Computer Ethics*. New Jersey: Prentice-Hall, 1985.
- Kurzweil, R. *The Age of Spiritual Machines*. New York: Penguin, 1999.
- Lenk, H., Ropohl, G. (eds.). *Technik und Ethik*. Stuttgart: Reclam, 1987.
- Karafyllis, N.C. *Versuch über den Menschen zwischen Artefakt und Lebewesen*. Paderborn: mentis, 2003.
- Lodge, J., Nagel, D. (eds.) *Ethical Issues of Networked Toys*. In: *International Review of Information Ethics*, Vol. 27, 2018. http://www.i-r-i-e.net/current_issue.htm
- McCorduck, P. (1979). *Machines Who Think. A Personal Inquiry into the History and Prospects of Artificial Intelligence*. San Francisco: Freeman.
- McLuhan, M. (1962). *The Gutenberg Galaxy: the making of typographic man*. Toronto, Canada: University of Toronto Press.
- Marx, K. (1867). *Das Kapital. Kritik der politischen Ökonomie. Marx-Engels-Gesamtausgabe (MEGA), 1. Band*. Berlin: de Gruyter. <http://telota.bbaw.de/mega/> (Engl. quote from *Capital. A Critique of Political Economy*, Transl. S. Moore, E. Aveling (2015). www.marxists.org/archive/marx/works/download/pdf/Capital-Volume-I.pdf).
- Moravec, H. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, Mass.: Harvard University Press, 1988.
- Nakada, M. *Ethical and Critical Analysis of the Meanings of 'Autonomy' of Robot Technology in the West and Japan: Themes Underlying Discussions of the Ethical Aspects of Robotics and Human-Robot Interaction*. In: T. Kimura, M. Nakada, K. Suzuki, Y. Sankai (Eds): *Cybernetics Technical Reports. Special Issue on Roboethics*. University of Tsukuba, CYB-2011-001 - CYB-2011-008, 2011, pp. 6-91.

- Nakada, M., Capurro, R. *An Intercultural Dialogue on Roboethics*. In: M. Nakada, R. Capurro (eds.): *The Quest for Information Ethics and Roboethics in East and West. Research Report on trends in information ethics and roboethics in Japan and the West*. ReGIS (Research Group on the Information Society) and ICIE (International Center for Information Ethics), ISSN 2187-6061, 2013, pp. 13-22.
- Nakada, M., Capurro, R., Sato, K. (eds.) *Critical Review of Information Ethics and Roboethics in East and West. Master's and Doctoral Program in International and Advanced Japanese Studies, Research Group for "Ethics and Technology in the Information Era"*, University of Tsukuba, ISSN 2432-5414, 2017.
- Nagenborg, M., Capurro, R. *Ethical Evaluation. Part of D.3.2 (Evaluation Report)*, 2012. <http://www.etica-project.eu/deliverable-files>
- Nagenborg, M., Capurro, R., Weber, J., Pingel, Chr. *Ethical Regulations on Robotics in Europe*. In: *AI & Society*, 22, 2008, pp. 349-366.
- Negrotti, M. *Naturoids. On the Nature of the Artificial*. New Jersey: World Scientific, 2002.
- Negrotti, M. *The Reality of the Artificial. Nature, Technology and Naturoids*. Heidelberg: Springer, 2012.
- Negrotti, M. *The Theory of the Artificial*. Exeter, England: intellect, 1999.
- Orwell, G. *Animal Farm. A Fairy Story*. London: Penguin Books, 1987.
- Pascal, B. *Pensées*. Paris: Gallimard, 1977.
- Rahner, K. *Geist in Welt (Spirit in the World)*. München: Kösel (2nd ed.), 1957.
- Rodotà, S. *Il nuovo Habeas Corpus: La persona costituzionalizzata e la sua autodeterminazione (The New Habeas Corpus: The New Constitutionalised Person and its self-determination)*. In: S. Rodotà, M. Tallachini (eds.): *Ambito e fonti del biodiritto*. Milano: Giuffrè, 2010, pp. 169-230.
- Samek, Toni, and Lynette Shultz, eds. *Information Ethics, Globalization and Citizenship: Essays on Ideas to Praxis*. McFarland, 2017.
- Searle, J.R. *Minds, Brains, and Programs*. In: *Behavioral and Brain Sciences* 3 (3), 1980, pp. 417-457.
- Simon, H. *The Sciences of the Artificial*. Cambridge, Mass.: The M.I.T. Press, 1969.
- Sloman, A. *The Computer Revolution in Philosophy. Philosophy, Science and Models of Mind*. The University of Sussex: The Harvester Press, 1978.
- Stahl, B.C. *Ethics of European Institutions as Normative Foundation of Responsible Research and Innovation in ICT*. In: M. Kelly, J. Bielby (eds.): *Information Cultures in the Digital Age. A Festschrift in Honor of Rafael Capurro*. Heidelberg: Springer, 2016, pp. 207-219.
- Stengel, L., Nagenborg, M. *Reconstructing European Ethics. How does a Technology become an Ethical Issue at the Level of the EU? ETICA, Deliverable 3.2.2 Annex I*, 2011. http://www.capurro.de/etica_deliv3.2.2.annex.pdf
- Turing, A. M. *Computing Machinery and Intelligence*. In: *Mind*, Vol. LIX, Issue 235, 1950, pp. 433-460
- Turing, A. M. *Intelligent Machinery, report written for the National Physical Laboratory*. <http://www.turingarchive.org/viewer/?id=127&title=1> Accessed Jan 27, 2019

Tzafestas, S.G. Roboethics. A Navigating Overview. Heidelberg: Springer, 2016.

VDA (Verein der Automobilindustrie) Dialogreihe Mobilität von morgen. Gespräch mit dem Vorstandsvorsitzenden der Continental AG, Dr. Elmar Degenhart, 2016. <https://www.mobilitaet-von-morgen.de/6-dialog/wir-brauchen-klare-ethische-regeln>

Wallach, W., Allen, C. Moral Machines: Teaching Robots Right from Wrong. Oxford University Press, 2009.

Weinberg, A. E. Science, Government, and Information. A report of The President's Science Advisory Committee, 1963. <http://garfield.library.upenn.edu/papers/weinbergreport1963.pdf>

Weizenbaum, J. ELIZA — A Computer Program for the Study of Natural Language Communication between Man and Machine. In: Communications of the Association for Computing Machinery 9, 1966, pp. 36-45.

Weizenbaum, J. Computer Power and Human Reason: From Judgement To Calculation. San Francisco: W.H. Freeman, 1976.

Wiener, N. Cybernetics: or Control and Communication in the Animal and the Machine. Cambridge, Massachusetts: The MIT Press (2nd ed.), 1965.

Wiener, N. The Human Use of Human Beings. Cybernetics and Society. London: Free Ass. Books, 1989.

Winograd, T., Flores, F. Understanding Computers and Cognition. A New Foundation for Design. New Jersey: Ablex, 1986.