Müller, Vincent C. (forthcoming), 'History of digital ethics', in Carissa Véliz (ed.), *Oxford Handbook of Digital Ethics* (Oxford: Oxford University Press).

History of digital ethics

Vincent C. Müller

TU Eindhoven (U Leeds, Alan Turing Institute)
www.sophia.de

Late Draft, 21st May, 2021

Abstract: Digital ethics, also known as computer ethics or information ethics, is now a lively field that draws a lot of attention, but how did it come about and what were the developments that lead to its existence? What are the traditions, the concerns, the technological and social developments that pushed digital ethics? How did ethical issues change with digitalisation of human life? How did the traditional discipline of philosophy respond? The article provides an overview, proposing historical epochs: 'pre-modernity' prior to digital computation over data, via the 'modernity' of digital data processing to our present 'post-modernity' when not only the data is digital, but our lives themselves are largely digital. In each section, the situation in technology and society is sketched, and then the developments in digital ethics are explained. Finally, a brief outlook is provided.

1. Introduction

The history of digital *ethics* as a field followed the development and use of digital *technologies* in society, and it often mirrors the ethical concerns of the pre-digital technologies that were replaced – it is only fairly recently that digital technologies have posed questions that are truly new. When 'data processing' became a more common activity in industry and public administration in the 1960s, the concerns of ethicists were known issues like *privacy, data security* and *power* through information access. Today, digital ethics involves old issues that took on a new quality due to digital technology, such as *surveillance, news,* or *dating*; but it also covers new issues that did not exist at all, such as *automated weapons, search engines, automated decision-making,* and *existential risk from AI*.

The terms used to name the expanding discipline have also changed over time: We started with 'computer ethics' (Bynum 2001; Johnson 1985; Vacura 2015), then more abstract terms like 'information ethics' were proposed, and now some use a new term: 'digital ethics' (Capurro 2010), as this handbook does. We also have digital ethics for particular areas, such as 'ethics of AI', 'data ethics', 'robot ethics', etc.

There are reasons for these changes: 'computer ethics' now sounds dated because it focuses attention on the machines, which made good sense when they were visible big boxes, but began to make less sense when many technical devices invisibly included computing. The more ambitious notion of 'information ethics' involves a digital ontology (Capurro 2006) and faces a significant challenge to explain the role of the notion of 'information'; see (Floridi 1999) vs. (Floridi and Taddeo 2016). Also, the term 'information ethics' is sometimes considered in contexts in which information is not computed, e.g. in 'library and information science'. Occasionally one hears 'cyberethics' (Spinello 2020), specifically dealing with the connected 'cyberspace' – probably now an outdated term, at least outside the military. In this confusion, some people use 'digital' as the new term, which captures most relevant phenomena and moves away from the machinery to their use – as does this handbook. One might argue that the process of 'computing' is still fundamental, but that we will probably soon care less whether a device uses computing (analogue or digital) – rather like we don't care much which energy source the engine in a car uses. The notion of 'data' will continue to make sense, but in the future, I suspect that terms like 'computing' and 'digital' will just merge into 'technology'.

Given that this handbook has articles on the current state of the art, this article on the history of the field will not attempt to say much about the present. Instead, it tries to give a historical context to the current debates, both in debates during the early days of information technology (IT), from the 1940s to the 1970s, when IT was an expensive technology available only in well-funded central 'computation centres'; then roughly the 1980s to the early 2000s, with networked personal computers entering offices and households; finally the last 15 years or so with 'smart' phones and other 'smart' devices being used privately – for new purposes that emerge with the devices.

This article is structured by two ideas, namely that a) technology drives ethics, and b) many issues that are now in 'digital ethics' predate digital technology. There is a certain tension between these two statements: The question is when 'technology drives ethics' and when that 'drive' is specific to 'digital' (computing)

technology. Since we think that b) is true, we must start before the invention of digital technology; in fact, even before the invention of writing.

We propose to divide the history into three main sections: pre-modernity (before the invention digital technology), modernity (with digital technology, but analogue lives), post-modernity (with digital technology and digital lives). We hope that this organisation matches the social developments of these periods, but we make no claim that the terminology used here is congruent with a standard history of digital society. In each section, we will briefly look at the technology, and then at digital ethics. Finally, it may be mentioned that there are significant research desiderata in the field; a detailed history of digital ethics, and indeed of applied or practical ethics, is yet to be written.

2. Pre-Modernity: Talking and Writing

2.1. Technology & Society

A fair amount of the concerns of information ethics is about privacy, information security, power through information, etc. These issues existed well before the computing age. They do not even require that information is represented in symbolic form – they also feature in village gossip.

One significant step for this timeline, however, were the beginnings of symbols and iconic representations from cave-paintings onwards (cf. Sassoon and Gaur 1997). These allow to maintain records that do not immediately vanish, like speech does, and some of which can be transported to another place. It may be useful to differentiate (1) representation *for* someone, or *intentional representation*, and (2) *representation per se*, when something represents something else because that is its function in a system (assuming this is possible without intentional states). The word 'tree', pronounced by someone, is an intentional representation (type 1); the non-linguistic representation of a tree in the brain of an organism that sees the tree is a non-intentional representation (type 2) (Müller 2007). Evidently, one major step that is relevant for digital ethics was the invention and use of *writing* – for the representation of natural language but also for mathematics and other purposes. Symbols in writing are digital, i.e. they have a sharp boundary with no intermediate stages (something is either an 'A' or a 'B', it cannot be a bit of both) and they are perfectly reproducible, i.e. one can write the exact same word or sentence more than once. The replication of writing and images, in print, also multiplies the impact that goes with that writing

– what is printed can be transported, remembered, and read by many people. It can become part of the cultural heritage.

A further major step is the transmission of speech and symbols over large distances and then to larger audiences through telegraph, mail, radio and TV. Suddenly, a single person speaking could be heard and even seen by millions of others around the globe, even in real time.

2.2. Ethics

There is a significant body of ethical and legal discussion on pre-digital information handling, especially after the invention of writing, printing and mass communication. Much of it is still the law today, such as the privacy of letters and other written communication, the press laws and laws on libel. The privacy of letters was legally protected in the early days of postal services in the early 18th Century, e.g. in the "Prussian New Postal Order" of 1712, (Matthias 1812: 54). Remarkably, several of these laws have lost their teeth in the digital era, e.g. email is often not protected by the privacy of letters, and online publications are often not covered by press law.

The central issue of privacy, often connected with 'data protection' started around 1900 (Warren and Brandeis 1890), developed into a field (Hoffman 1973; Martin 1973; Westin 1968) and is still a central discussion today; from classical surveillance (Macnish 2017) governance (Bennett and Raab 2003) and ethical analysis (Roessler 2017; van den Hoven et al. 2020) to analysis for activism (Véliz 2020). The very close link between ethics and law, or indeed social science, has been somewhat lost since, and we are only recently re-establishing it, now that law must related to ethics, and ethics must relate to societal developments. The power of information and misinformation was well understood after the invention of printing, but especially after the invention of mass media like radio and TV and their use in propaganda – media studies and media ethics became standard fields after the Second World War. Media ethics is still an important aspect of digital ethics (Ess 2014), especially the aspect of the 'public sphere' (Habermas 1962).

Apart from this tradition of more 'societal' ethics, there is a more personal ethics of *professional responsibility* that started in this area – and had impact in the digital era. The influential *Institute of Electrical and Electronics Engineers* (IEEE, initially AIEE) adopted its first "Principles of professional conduct for the guidance of the electrical engineer" in 1912 (AIEE 1912). 'Engineering Ethics' is thus older than ethics of computing – but, interestingly, the electrical and telephone industries in the USA managed to get an exception to the demand that engineers hold a

professional license (PE). This move may have had far-reaching impact into the computer science of today, which usually does not see itself as a discipline of engineering, and bound by the ethos of engineers – though there are computer scientists that would want to achieve the recognition as a profession and thus the ethos of 'being a good engineer' (in many countries, engineering has high status and computer science degrees are 'diplomas in engineering').

Up to this point, we see the main ethical themes of privacy and data security, power of information, and professional responsibility.

3. Modernity: Digital ethics in IT

3.1. Technology & Society

As a rough starting point for this part of the timeline one should take the first design for a universal computer with Babbage's 'Analytic Engine' in about 1840; the first actual universal computer was feasible only when computers could use electronic parts, starting with Zuse's Z3 in 1941, followed by the independently developed ENIAC in 1945, and Manchester Mark I in 1949 and then many more machines, mostly due to military funding during (Ifrah 1981). All major computers since then have been electronic universal digital computers with stored program. Shortly after WWII, we have the beginnings of the science of 'Informatics' with 'Cybernetics' (Ashby 1956; Wiener 1948) and C. E. Shannon's "A Mathematical Theory of Communication" (Shannon 1948). In 1956, J. McCarthy, M. L. Minsky, N. Rochester and C. E. Shannon organised the Dartmouth conference on 'Artificial Intelligence', thus coining the term (McCarthy et al. 1955). Less than 10 years later, H. Simon predicts "Machines will be capable, within 20 years, of doing any work that a man can do." (Simon 1965: 96). In 1971, integrated processor (microprocessor) computers started, with all integrated circuits in one microchip. This is effectively the modern computer area that made 'personal computers' possible in the 1980s. Up to that point, computers were big and very expensive devices, only used by large corporations, research centres or public entities for 'data processing'.

Ray Kurzweil has put the development from WWII to the present with characteristic panache:

> Computers started out as large remote machines in air-conditioned rooms tended by white coated technicians. Subsequently they moved onto our desks, then under our arms, and now in our pockets. Soon, we'll routinely put them inside our bodies and brains. Ultimately we will become more nonbiologial than biological. (Kurzweil 2002).

3.2. Ethics

3.2.1. Professional ethics

The first discussions about ethics and computers were about the ethics of the people who work professionally in computing – what they should or should not do. In that phase, a computer scientist was an expert, rather like a doctor or a mechanical engineer, and the question arose whether the new 'profession' needed an ethics. These early discussions of computer ethics often had a certain tinge of moralising, of having discovered an area of life that had escaped the attention of ethicists so far, but where immorality looms. In those days it is rare to find the more positive approach that practitioners face ethical problems that expert analysis might help to resolve; what one does find is the concern that a particular technology may have negative effects on society. This suspicion of immorality was often supported by the innocent view of practitioners that technology is neutral and our aims laudable, thus an 'ethics of technology' is not needed – a view one finds even today.

The early attempts at a professional ethics moved into Computer Science pretty much at the beginning of the discipline, e.g. the US *Association for Computing Machinery* (ACM) adopted "Guidelines for Professional Conduct in Information Processing" on 11.11.1966 (Parker 1968), and Donn Parker pushed this agenda in his discipline in the ensuing years. The current version is called the "ACM Code of Ethics and Professional Conduct" (ACM 2018).

3.2.2. Responsible technology

The use of nuclear (atomic) bombs in the Second World War and the discussion about the risk of generating electricity in nuclear power stations (from the late 1950s) fueled an increasing concern with the limits of technology in the 1960s. This political development is closely connected to the political developments in 'the generation of 1968' on the political left in Europe and the United States. The 'Club of Rome' was and is a group of high-level politicians, scientists, and industry-leaders that deals with basic long-term problems of humankind. In 1972, it published the highly influential book *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind* (Club of Rome 1972). It argued that the industrialised world was on an unsustainable trajectory of economic growth, using up finite resources (e.g. oil, minerals, farmable land), and increasing pollution, on the background of an increasing world population.

This book and other similar discussions fueled a generally more critical view of technology and the growth it enables. They lead to a field of 'technology

assessment' in terms of longer term impacts that has also dealt with information technologies (Grunwald 2002). This area of the social sciences is influential in political consulting and has some academic institutes (e.g. at the Karlsruhe Institute of Technology). At the same time, a more political angle of technology is taken in the field of 'Science and Technology Studies' (STS), which is now a sizable academic field with programs, journals, and conferences. As books like *The Ethics of Invention* (Jasanoff 2016) show, concerns in STS are often quite similar to those in ethics, though typically with a more 'critical' and more empirical approach. These STS approaches have remained oddly separate from ethics of computing.

Concerns about *sustainable development*, especially 'the environment' have been prominent on the political agenda for about 40 years and they are now quite officially a central policy aim. In 2015, the UN adopted the "2030 Agenda for Sustainable Development" (United Nations 2015). Its 17 "Sustainable Development Goals" are now heavily influential, e.g. they guide the current development of official policy on AI. The 17 goals are, in brief: (1) No Poverty, (2) Zero Hunger, (3) Good Health and Well-being, (4) Quality Education, (5) Gender Equality, (6) Clean Water and Sanitation, (7) Affordable and Clean Energy, (8) Decent Work and Economic Growth, (9) Industry, Innovation and Infrastructure, (10) Reducing Inequality, (11) Sustainable Cities and Communities, (12) Responsible Consumption and Production, (13) Climate Action, (14) Life Below Water, (15) Life On Land, (16) Peace, Justice, and Strong Institutions, (17) Partnerships for the Goals.

3.2.3. Control

It had also been understood by some that science and engineering, generally, pose ethical problems. The prominent physicist, C. F .v. Weizsäcker predicted in 1968 that computer technology will fundamentally transform our lives in the coming decades (Weizsäcker 1968) and asked how we will have individual freedom in such a world, "i.e. freedom from the control of anonymous powers" (439). At the end of his article, he demands a Hippocratic oath for scientists. Soon after, in 1970, Weizsäcker became founding Director of the famous *Max Planck Institute for Research into the Life in a Scientific-Technical World*, co-directed by Jürgen Habermas since 1971. Even at that time, there was clearly a sense with major state funders that these issues deserved their own research institute.

The ACM already had a Special Interest Group 'Computers & Society' (SIGCAS) since 1969 – it is still a significant actor today and has published a newsletter/journal *Computers and Society* since 1972.

Norbert Wiener had warned of AI, even before the term was coined, e.g. in *Cybernetics*, he wrote:

> … we are already in a position to construct artificial machines of almost any degree of elaborateness of performance. Long before Nagasaki and the public awareness of the atomic bomb, it had occurred to me that we were here in the presence of another social potentiality of unheard-of importance for good and for evil. (Wiener 1948: 28).

Note the link to the atomic bomb, a starting point for the critical view on technology. In his later book *The Human Use of Human Beings* he warns of manipulation:

> … such machines, though helpless by themselves, may be used by a human being or a block of human beings to increase their control over the rest of the race or that political leaders may attempt to control their populations by means not of machines themselves but through political techniques as narrow and indifferent to human possibility as if they had, in fact, been conceived mechanically. (Wiener 1950)

For other aspects of Wiener's ethics, see (Bynum 2008: 26-30; 2015). Thus, in this phase, professional responsibility gains prominence as an issue, the notion of *control* through information and machinery comes up as a theme, and there is a general concern about the longer-term impacts of technology – a concern that shapes much of politics today.

4. Post-Modernity

4.1. Technology & Society

In this part of the quick timeline, from 1980 to today (2021), I will use a typical student in a wealthy country, like the United Kingdom, as an illustration. I think this timeline is useful because it is easy to forget how the availability and use of computers has changed in the last decades, and even the last years. (If this text is read a few years after writing, it will seem quaintly old-fashioned.) We will see that this is the phase in which computers enter peoples' lives and digital ethics becomes a discipline.

In the first half of the 1980s, a student would have seen a 'personal computer' (PC) in a business context, and towards the end of the 1980s they would probably own one. These PCs were not networked, unless on university premises, so data exchange was through floppy disks. Floppy disks held 360KB, later 720 KB and 1.44 MB; if the PC had a hard drive at all, it would hold ca. 20-120 MB. After 1990,

if private PCs had network connections, that would be through modem dial-in on analogue telephone lines that would mainly serve links to others in the same network (e.g. CompuServe or AOL), allowing email and ftp (file-transfer protocol). Around the same time, PCs moved from a command-line to a graphic interface, e.g. MS Windows, Mac OS or UNIX. Students would use electrical typewriters or university-owned computers for their writing until well into the 1990s. The first WWW page came online in 1990; institutional web pages became common in the late 1990s; around the same time a dial-in Internet connection at home through a modem became affordable, and Google was founded (1998). After 2000, it became common for a student to have a computer at home, with an Internet connection, though file-exchanges would still be mostly via physical data-carriers. By ca. 2010 the Internet connection would be 'always on' and fast enough for frequent use of www pages, and video; by ca. 2019 it would be fully digital (ISDN, ASDL, …) and its files would often be stored in 'cloud' spaces somewhere on the Internet; fibre-optic lines started to be used around 2020. With the COVID pandemic 2020-21, cooperative work online through live video became common.

Mobile phones (cell phones) became commonly affordable by students in the late 1990s, but these were just phones, increasingly miniaturised. The first 'smart' phone, the iPhone, was introduced in 2007. Around 2015, a student would own such a smart phone and would use that phone mostly for things other than calls; essentially as a portable tablet computer with Wi-Fi capability (but it would be called a 'phone', not a 'computer'). After 2015, the typical smart phone would be connected to the Internet at all times (with 3G). The frequent use of the WWW over phone Internet became affordable around 2018/19 (with 4G), at which time video calls and online teaching became possible and useful.

Together with smartphones, we now (2021) also begin to have other 'smart' devices that incorporate computers and are connected to the Internet (soon with 5G), especially portables, TVs, cars and homes – also known as the 'Internet of Things' (IoT). 'Smart' superstructures like grids, cities, and roads are developing as well. Sensors with digital output are becoming ubiquitous. In addition, a large part of our lives is digital (and thus does not need to be captured by sensors), and much of it conducted through commercial platforms and 'social media' systems. All these developments enable a surveillance economy.

While a 'computer' was easily recognised as a physical box until ca. 2010, it is now incorporated in a host of devices and systems, and often not perceived as such; perhaps even designed not to be noticed (e.g. in order to collect data). Much of computing has become a transparent technology in our daily lives: We use it

without special learning, do not notice its existence, or that computing takes place: "The most profound technologies are those that disappear" (Weiser 1991: 94).

For the purposes of digital ethics, the crucial developments of our student were the move from computers 'somewhere else' to her own PC (ca. 1990), the use of the WWW (ca. 1995) and her smartphone (ca. 2015); the current development is the move to computing as a 'transparent technology'.

4.2. Ethics

4.2.1. Establishment

The first phase of digital ethics, or computer ethics, was the effort in the 1980s and 90s to establish that there *is* such a thing, or that there *should be* such a thing – both within philosophy or applied ethics, and within computer science, especially the curriculum of computer science at universities. This 'establishment' is of significant importance for the academic field, since, once 'ethics' is an established component of degrees in computer science and related disciplines, there is a labour market for academic teachers, a demand for writing textbooks and articles, etc. (Bynum 2010). It is not an accident that the field was established beyond 'professional ethics' and general societal concerns around the same time as the move of computers from labs to offices and homes occurred.

The first use of 'computer ethics' was probably by Deborah Johnson in her paper "Computer ethics: New study area for engineering science students", where she remarked "Computer professionals are beginning to look toward codes of ethics and legislation to control the use of software" (Johnson 1978). Sometimes, (Bynum 2001) it is Walter Maner who credited with the first use for "ethical problems aggravated, transformed or created by computer technology" (Maner 1980). Again, professional ethics seems to have been the forerunner for computer ethics, generally.

A few years later, with fundamental publications like James H. [Jim] Moor's "What is computer ethics?" (Moor 1985), the first textbook (Johnson 1985), and three anthologies with established publishers (Blackwell, MIT Press, Columbia UP) one can speak of an established small discipline (Moor and Bynum 2002). These two texts by Moor and Johnson are still the most cited works in the discipline, together with classic texts on privacy, such as (Warren and Brandeis 1890) and (Westin 1968). As (Tavani 1999) shows, there is a steady flow of monographs, textbooks and anthologies in the 15 years that followed. In the 1990s, 'ethics' started to gain a place in many computer science curricula, thus generating demand for qualified faculty and for teaching material.

In terms of *themes*, we have the classical ones (privacy, information power, professional ethics, impact of technology) and we now have an increasing confidence that there is 'something unique' here. Maner says: "I have tried to show that there are issues and problems that are unique to computer ethics. For all of these issues, there was an essential involvement of computing technology. Except for this technology, these issues would not have arisen, or would not have arisen in their highly altered form." (Maner 1996).

In this vein, we now get a wider notion that includes issues that only come up in ethics of *robotics and AI*, e.g. manipulation, automated decision-making, transparency, bias, autonomous systems, existential risk, etc. (Müller 2020). More radically, digital ethics now covers the human *digital life*, online and with computing devices – both on an individual level and as a society, e.g. social networks (Vallor 2016). As a result, this handbook includes themes like human-robot interaction, online interaction, fake news, online relationships, advisory systems, transparency & explainability, discrimination, nudging, cybersecurity, and existential risk – in other words, the digital life is prominently discussed here; something that would not have happened even five years ago.

4.2.2. Institutional

The journal *Metaphilosophy* first published articles on computer ethics in the mid-1980s; T.W. Bynum and R. Reese had founded it in 1970. The more prominent journal *Minds and Machines,* founded by James Fetzer 1991, started publishing ethics papers under the editorship of James H. Moor (2001-10). The conferences series ETHICOMP (1995) and CEPE (1997) started in Europe, and specialised journals were established: *Journal of Information Ethics* (1992), *Science and Engineering Ethics* (1995), *Ethics and Information Technology* (1999), *Philosophy & Technology* (2010). The conferences on 'Computing and Philosophy' (CAP), since 1986 in North America, later in Europe and Asia, united to IACAP in 2011, increasingly have a strong division on ethical issues; as does the AISB (in the UK).

Within the academic field of philosophy, applied ethics and digital ethics have remained firmly marginal or specialist even now, with very few presentations at mainstream conferences, publications in mainstream journals, or posts at mainstream departments. As far as I can tell, no paper on digital ethics has appeared in places like *The Journal of Philosophy, Mind, Philosophical Review, Philosophy & Public Affairs* or *Ethics* to this day – while, significantly, there are papers in *Science, Nature* or *Artificial Intelligence*. Practically oriented fields are treated largely as the poor and slightly embarrassing cousin who has to work for a living, rather than having old money in the bank. What counts as 'a problem' is

still mostly defined through the tradition of philosophy, rather than permitting a problem to enter philosophy from the outside life. It is only very recently that some in these 'practical' fields have the ambition to have a real influence in traditional philosophy, and I would venture that this influence will be strong in the decades to come. It is interesting to note that the citation counts of academics in computing ethics and theory have surpassed those of comparable philosophers in related traditional areas, and similar trends are happening now with journals. One data point: As of 2019, the average article in *Mind* is cited 2 times within 4 years, while the average article in *Minds and Machines* is cited 3 times within 4 years – the number for the latter journal doubled in 3 years (https://www.scimagojr.com).

Several prominent philosophers have worked on theoretical issues around AI and computing (e.g. Dennett, Dreyfus, Fodor, Haugeland, Searle), given the more mainstream attachment to 'philosophy of cognitive science' and 'philosophy of mind' – and typically with a foundation of their careers in other areas of philosophy. This also applies to Jim Moor, who was one of the first people in digital ethics to hold a professorship at a reputed general university (Dartmouth College). In Europe, several technical universities had professors working in digital ethics early on, notably the TUs in the Netherlands, who founded a 4TU Centre for Ethics and Technology in 2007, joining Delft, Eindhoven, Twente and Wageningen. In the last decade, Floridi and Bostrom were appointed to professorships at Oxford, at the OII and FHI institutes. Coeckelbergh was appointed to a chair at the philosophy department in Vienna in 2015 (where Hrachovec had been active already). A few more people were and are active in philosophical issues of 'new media', e.g. Ch. Ess who moved to Oslo in 2012. The ethics of AI became a field only quite recently, with the first conference in 2012 (AGI-Impacts), but it now has its own institutes at many mainstream universities.

In other words, only five years ago, almost all scholars in digital ethics were at marginal institutions to mainstream philosophy. It is only in those last couple of years that digital ethics is becoming mainstream; many more jobs are advertised, senior positions are available to people in the field, younger faculty are picking up, and more established faculty at established institutions begin to deem these matters worthy of their attention.

We expect that mainstream philosophy will quickly pick up digital ethics in the coming years – the subject has shown itself to be mature, fruitful for classical philosophical issues, there is an obvious societal demand, and there are significant funding opportunities. Probably there is also some hype already. In the classic

notion of a 'hype cycle' for the expectations from a new technology, the development is supposed to go through several phases: After its beginnings at the 'Technology Trigger', it gains more and more attention, reaching a 'Peak of Inflated Expectations', after that a more critical evaluation begins and the expectations go down, reaching a 'Trough of Disillusionment'. From there, a realistic evaluation shows that there is some use, so we get the 'Slope of Enlightenment' and eventually the technology settles on a 'Plateau of Productivity' and becomes mainstream. The *Gartner Hype Cycle for AI, 2019* (Goasduff 2019) sees digital ethics itself at the 'peak of inflated expectations' … meaning it is downhill from here, for some time, until we hopefully reach the 'plateau of productivity'. (My own view is that this is wrong, since we see the beginnings of AI policy and stronger digital ethics now.)

4.2.3. Future

The state of the art at the present and an outlook into the future are given in the chapters of this handbook. Moor saw a bright future 20 years ago already: "The future of computer ethics: You ain't seen nothin' yet!" (Moor 2001), and he followed up with a programmatic plea for 'machine ethics' (Moor 2006). Moor opens his article with the bold statement:

> Computer ethics is a growth area. My prediction is that ethical problems generated by computers and information technology in general will abound for the foreseeable future. Moreover, we will continue to regard these issues as problems of computer ethics even though the ubiquitous computing devices themselves may tend to disappear into our clothing, our walls, our vehicles, our appliances, and ourselves. (Moor 2001: 89)

This prediction has undoubtedly held up until now. The ethics of the design and use of computers is clearly an area of very high societal importance and we would do well to catch problems early on – this is something we failed to do in the area of privacy (Véliz 2020) and some hope that we will do in the area of AI (Müller 2020).

However, as Moor mentions, there is also a very different possible line that was developed around the same time: Bynum reports on an unpublished talk by Deborah G. Johnson with the title "Computer Ethics in the 21st Century", at the 1999 ETHICOMP conference:

> On Johnson's view, as information technology becomes very commonplace – as it gets integrated and absorbed into our everyday surroundings and is perceived simply as an aspect of ordinary life – we may no longer notice its

> presence. At that point, we would no longer need a term like 'computer ethics' to single out a subset of ethical issues arising from the use of information technology. Computer technology would be absorbed into the fabric of life, and computer ethics would thus be effectively absorbed into ordinary ethics. (Bynum 2001: 111f) (cf. Johnson 2004)

On Johnson's view, we will have applied ethics and the ethics will concern most themes, such as 'information privacy' or 'how to behave in a romantic relationship' – and much of this will be taking places with or through computing devices, but it will not matter (even though many things will remain that cannot be done without such devices). In other words, the 'drive' of technology we have seen in this history will come to a close, and the technology will become transparent. This transparency will likely have ethical problems itself – it enables surveillance and manipulation. If Johnson is right, however, we will soon have the situation that all too much is digital and transparent, and thus digital ethics is in danger of disappearing. In Molière's play, this bourgeois who wants to become a gentleman tells his 'philosophy master':

> "Oh dear! For more than forty years I have been speaking prose while knowing nothing of it, and I am most obliged to you for telling me so."
>
> Molière, *Le Bourgeois gentilhomme* (Act II) 1670

5. Conclusion, questions

One feature that is characteristic of the new developments in digital ethics, and in applied philosophy, generally, is how a problem becomes a problem worth investigating. In traditional philosophy, the criterion is often that there is already a discussion in the tradition and that there is something philosophically *interesting* about it, something unresolved – and typically we do not need to ask again whether that problem is really worth discussing, or whether it perhaps relies on assumptions we should not make (so we will get people who seriously ask whether Leibniz or Locke was right on the origin of ideas, for example). In digital ethics, what counts as a problem also includes the demand to philosophically *interesting*, but more importantly, whether it has *relevance*. Quite often this means that the problem first surfaces in fields other than philosophy. The initially dominant approach of *professional ethics* had a touch of 'policing' about it, of checking that everyone behaves - that moralising gives ethics a bad name and it typically comes too late. More modern digital ethics tries to make people sensitive in the design process ('ethics by design') and to pick up problems

where people really do not know what the ethically right thing to do is – these are the proper ethical problems that deserve our attention.

For the relation of ethics and computer ethics, Moor seemed right in this prediction:

> The development of ethical theory in computer ethics is sometimes overstated and sometimes understated. The overstatement suggests that computer ethics will produce a new ethical theory quite apart from traditional ethical notions. The understatement suggests that computer ethics will disappear into ordinary ethics. The truth, I predict, will be found in the middle. […] My prediction is that ethical theory in the future will be recognizable but reconfigured because of work done in computer ethics during the coming century. (Moor 2001: 91)

As philosophers, we must do more than *export* an expertise from philosophy or ethics to practical problems, we must also *import* insights from these debates back to philosophy. The discipline can feed largely on the societal demand and the real impact philosophical insights can have in this area, but in order to secure its place within philosophy, we must show that the work is both technically serious and has real potential to shed light on traditional issues. It seems obvious that this is the case. Consider the question of when an artificial agent truly *is* an agent that is responsible for their actions – that discussion seems to provide a new angle to these debates that traditionally focused on human beings or animals. Now we can set the conceptual question anew and also provide evidence from experiments with *making* things, rather than from passive observation.

Nearly 250 years ago, Immanuel Kant stated that our philosophical reasoning is about four main questions: "1. What can I know? 2. What should I do? 3. What can I hope for? 4. What is the human?" (Kant 1800: 26), (1-3 already in Kant 1781: A805 & B833). The philosophical reflection on digital technology contributes to all four of these.

Acknowledgements
I am grateful for useful comments to Guido Löhr, Karsten Weber and Eleftheria Deltsou, and for detailed reviewing to Carissa Véliz, Maximilian Karge and Jeff White.

References

ACM (2018), 'ACM Code of Ethics and Professional Conduct'. <https://ethics.acm.org>.

AIEE (1912), 'Principles of professional conduct for the guidance of the electrical engineer', *Transactions of the American Institute of Electrical Engineers,* 31.

Ashby, W Ross (1956), *An Introduction to Cybernetics* (Eastford, CT: Martino Fine Books).

Bennett, Colin J and Raab, Charles (2003), *The governance of privacy: Policy instruments in global perspective* (3rd 2017 edn.; Cambridge, Mass.: MIT Press).

Bynum, Terrell W. (2001), 'Computer ethics: Its birth and its future', *Ethics and Information Technology,* 3 (2), 109-12.

— (2008), 'Milestones in the history of information and computer ethics', in Kenneth Einar Himma and Herman T Tavani (eds.), *The handbook of information and computer ethics* (New York: Wiley), 25-48.

— (2010), 'The historical roots of information and computer ethics', in Luciano Floridi (ed.), *The Cambridge handbook of information and computer ethics* (Cambridge: Cambridge University Press), 20-38. <https://www.cambridge.org/core/books/cambridge-handbook-of-information-and-computer-ethics/AA0E1E64AE997C80FABD3657FD8F6CA8>

— (2015), 'Computer and Information Ethics', *The Stanford Encyclopedia of Philosophy (Summer 2015 Edition)* (Stanford, Cal.: CLSI). <https://plato.stanford.edu/archives/sum2018/entries/ethics-computer/>

Capurro, Raphael (2006), 'Towards an ontological foundation of information ethics', *Ethics and Information Technology,* 8 (4), 175-86.

— (2010), 'Digitial ethics', in The Academy of Korean Studies (ed.), *Civilization and Peace* (The Academy of Korean Studies), 203-14. <http://www.capurro.de/korea.html>

Club of Rome (1972), *The Limits to Growth* (New York: Potomac Associates).

Ess, Charles (2014), *Digital media ethics* (2nd edn.; Cambridge: Polity Press).

Floridi, Luciano (1999), 'Information ethics: On the philosophical foundation of computer ethics', *Ethics and Information Technology,* 1 (1), 33-52.

Floridi, Luciano and Taddeo, Mariarosaria (2016), 'What is Data Ethics?', *Phil. Trans. R. Soc. A,* 374 (2083).

Goasduff, Laurence (2019), 'Top trends on the Gartner Hype Cycle for Artificial Intelligence, 2019', September 12, 2019. <https://www.gartner.com/smarterwithgartner/top-trends-on-the-gartner-hype-cycle-for-artificial-intelligence-2019/>.

Grunwald, Armin (2002), *Technikfolgenabschätzung – eine Einführung* (Berlin: Edition Sigma).

Habermas, Jürgen (1962), *Strukturwandel der Öffentlichkeit. Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft* (Neuwied/Berlin: Luchterhand).

Hoffman, Lance J. (1973), *Security and privacy in computer systems* (Los Angeles: Melville Publications).
Ifrah, Georges (1981), *Histoire universelle des chiffres* (Paris: Editions Seghers).
Jasanoff, Sheila (2016), *The ethics of invention: Technology and the human future* (New York: Norton).
Johnson, Deborah G (1978), 'Computer Ethics: New Study Area for Engineering Science Students', *Professional Engineer,* 48 (8), 32-4.
— (1985), *Computer ethics* (Englewood Cliffs (NJ): Prentice Hall).
— (2004), 'Computer Ethics', in Luciano Floridi (ed.), *The Blackwell Guide to the Philosophy of Computing and Information* (Oxford: Blackwell), 65-74.
Kant, Immanuel (1781), *Kritik der reinen Vernunft*, ed. Wolfgang Weischedel (A/B edn., Werkausgabe III & IV; Frankfurt: Suhrkamp 1956).
— (1800), *Logik*, ed. Wolfgang Weischedel (Werkausgabe VI; Frankfurt: Suhrkamp 1956).
Kurzweil, Ray (2002), 'We Are Becoming Cyborgs'. <http://www.kurzweilai.net/we-are-becoming-cyborgs>, accessed 13 June 2019.
Macnish, Kevin (2017), *The ethics of surveillance: An introduction* (London: Routledge).
Maner, Walter (1980), *Starter kit in computer ethics* (Hyde Park, New York: Helvetia Press and the National Information and Resource Center for Teaching Philosophy).
— (1996), 'Unique Ethical Problems in Information Technology', *Science and Engineering Ethics,* 2 (2), 137–54.
Martin, James (1973), *Security, accuracy, and privacy in computer systems* (Englewood Cliffs: Prentice-Hall).
Matthias, Wilhelm Heinrich (1812), *Darstellung des Postwesens in den königlich preußischen Staaten* (Berlin: Selbstverlag).
McCarthy, John; Minsky, Marvin; Rochester, Nathaniel and Shannon, Claude E. (1955), 'A proposal for the Dartmouth summer research project on artificial intelligence', <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>, accessed October 2006.
Moor, James H. (1985), 'What Is computer ethics?', *Metaphilosophy,* 16 (4), 266-75.
— (2001), 'The future of computer ethics: You ain't seen nothin' yet!', *Ethics and Information Technology,* 3 (2), 89-91.
— (2006), 'The nature, importance, and difficulty of machine ethics', *IEEE Intelligent Systems,* 21 (4), 18-21.
Moor, James H. and Bynum, Terrell Ward (2002), *Cyberphilosophy: The intersection of philosophy and computing* (Oxford: Blackwell).
Müller, Vincent C. (2007), 'Is there a future for AI without representation?', *Minds and Machines,* 17 (1), 101-15.
— (2020), 'Ethics of artificial intelligence and robotics', in Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Summer 2020; Palo Alto: CSLI, Stanford University), 1-70. <https://plato.stanford.edu/entries/ethics-ai/>

Parker, Donn B (1968), 'Rules of ethics in information processing', *Communicaitions of the ACM,* 11, 198-201.
Roessler, Beate (2017), 'Privacy as a human right', *Proceedings of the Aristotelian Society,* 2 (CXVII).
Sassoon, Rosemary and Gaur, Albertine (1997), *Signs, symbols and icons: Pre-history of the computer age* (Exeter: Intellect Books).
Shannon, Claude E. (1948), 'A mathematical theory of communication', *Bell Systems Technical Journal,* 27 (July, October), 379–423, 623–56.
Simon, Herbert (1965), *The shape of automation for men and management* (New York: Harper & Row).
Spinello, Richard A (2020), *Cyberethics: Morality and law in cyberspace* (Jones & Bartlett Learning).
Tavani, Herman T (1999), 'Computer ethics textbooks: a thirty-year retrospective', *ACM SIGCAS Computers and Society,* (September), 26-31.
United Nations (2015), 'The 2030 agenda for sustainable development'. <https://sustainabledevelopment.un.org/post2015/transformingourworld>.
Vacura, Miroslav (2015), 'The history of computer ethics and its future challenges', *Information technology and society interaction and independence (IDIMT 2015)* (Vienna), 325-33.
Vallor, Shannon (2016), 'Social Networking and Ethics', *The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta,* Summer 2016 Edition. <https://plato.stanford.edu/entries/ethics-social-networking/>.
van den Hoven, Jeroen; Blaauw, Martijn; Pieters, Wolter and Warnier, Martijn (2020), 'Privacy and Information Technology', *The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta,* Summer 2020 Edition. <https://plato.stanford.edu/archives/sum2020/entries/it-privacy/>.
Véliz, Carissa (2020), *Privacy is power* (London: Penguin).
Warren, Samuel D and Brandeis, Louis D (1890), 'The right to privacy', *Harvard Law Review,* 4 (5), 193-220.
Weiser, Mark (1991), 'The computer for the 21st Century', *Scientific American,* 265 (3), 94-104.
Weizsäcker, C. F. v. (1968), 'Die Wissenschaft als ethisches Problem', *Physikalische Blätter,* 10, 433-41.
Westin, Alan F (1968), 'Privacy and freedom', *Washington & Lee Law Review,* 25 (166).
Wiener, Norbert (1948), *Cybernetics: or control and communication in the animal and the machine* (1961 2nd. edn.; Cambridge, Mass.: MIT Press).
— (1950), *The Human Use of Human Beings* (Boston: Houghton Mifflin).