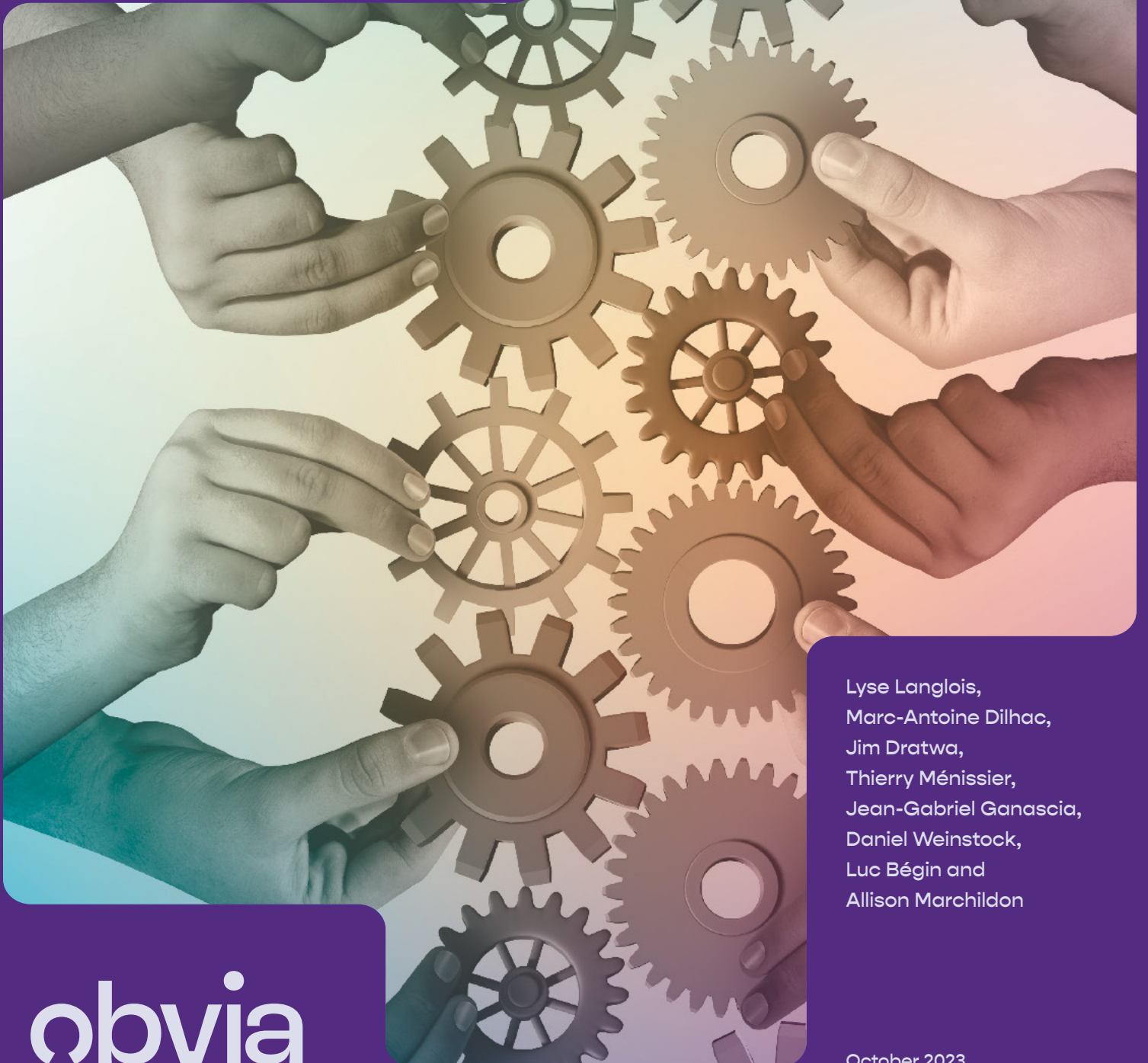


Ethics at the heart of AI



Lyse Langlois,
Marc-Antoine Dilhac,
Jim Dratwa,
Thierry Ménissier,
Jean-Gabriel Ganascia,
Daniel Weinstock,
Luc Bégin and
Allison Marchildon

obvia

October 2023

Foreword

This document is the result of a collaborative effort involving specialists in ethics, philosophy, computer science and economics. Its goal is to detail and clarify the role ethics should play in the age of Artificial Intelligence (AI) by highlighting how this notion can be applied and implemented effectively and successfully. It advocates for an ethic focused on reflexivity and dialogue, and it concerns all those involved in the development of AI, whether directly or indirectly. The document also highlights the practical methodological approach used to construct the Montreal Declaration, and also proposes a number of recommendations. In short, this paper argues for the inclusion of a genuine ethical reflection at all stages of the AI developmental process. It is a call for collaboration between ethicists, developers and members of the industry, to truly **put ethics at the heart of AI**.

Authors:

Lyse Langlois, Ph.D, Professor, Department of Industrial Relations, Université Laval, OBVIA CEO

Marc-Antoine Dilhac, Ph.D, Professor, Department of Philosophy, Université de Montréal, OBVIA Scientific Director-Governance and International Collaboration

Jim Dratwa, Ph.D, Advisor to the European Commission and Head of the Ethics in Science and New Technologies Team

Thierry Ménissier, Ph.D, Professor of Philosophy at Université Grenoble Alpes, Head of the Ethics and AI Chair at the Grenoble MIAI Institute (PIA 3IA)

Jean-Gabriel Ganascia, Ph.D, Professor of Computer Science at Sorbonne Université, Chairman of the CNRS Ethics Committee

Daniel Weinstock, Ph.D, Professor of Philosophy, Law Department, McGill University

Luc Bégin, Ph.D, Professor, Faculty of Philosophy, Université Laval

Allison Marchildon, Ph.D, Professor, Department of Philosophy and Applied Ethics, Université de Sherbrooke, and co-head of OBVIA's Ethics, Governance and Democracy Axis.

Coordination and research support :

Félix-Arnaud Morin-Bertrand, M.A,
Research professional at OBVIA

Acknowledgements :

Nathalie de Marcellis-Warin, Ph.D, Professor in the Department of Mathematics and Industrial Engineering at Polytechnique Montréal, CEO of CIRANO, OBVIA Scientific Director-Tools, Monitoring and Surveys

Translation and linguistic revision:

Michael Power and **Claire Lapointe**

Produced with the financial support of the Fonds de recherche du Québec

Québec 

Fonds de recherche – Nature et technologies
Fonds de recherche – Santé
Fonds de recherche – Société et culture

ISBN: 978-2-925138-27-3

Table of contents

Introduction	4
---------------------	---

Background	5
-------------------	---

1. Ethics and AI ethics	7
1.1. Ethics versus checklists	8
1.2. Ethical and technical standards	9

2. The ethics of AI	11
2.1. Ethics applied to AI	12
2.1.2 The origins of applied ethics: the case of bioethics	13
2.2. AI ethics versus “algorithmethics”	14

3. Methods of AI ethics	15
3.1. Deliberative ethics: the example of the Montreal Declaration	16

Conclusion	18
Recommandations	19

Lexicon	20
----------------	----

Bibliography	21
---------------------	----

Introduction

Nowadays, Artificial Intelligence (AI) occupies an unprecedented place in our lives. Indeed, AI is now integrated into multiple spheres of individual and social activities, influencing our choices, our relationships with others, and the way we work and learn. More broadly, digital technologies that underpin AI are changing the organisation of our world and, simultaneously, the way we understand and interact with it (Floridi, 2023). In recent years, ethics has been used to reflect on the concerns associated with the growing deployment of AI in society (Koniakou, 2023). The expanding interest in ethics has led to widely recognized ethical principles in AI, and even to the definition of an international standard (UNESCO, 2021). Despite this important step, **much remains to be done to demystify ethics and the benefits it can generate in terms of individual, organisational and strategic skills.** Indeed, the deployment of **generative artificial intelligence** models is currently raising issues in terms of education, law, labor, culture and democracy. In this context, ethics becomes essential as its vocation is eminently practical. Its place, function and application deserve to be seen as essential components in the deployment of a moral capability (Nussbaum, 2012) necessary to lead a dignified and free life, the purpose of which is human progress.

Ethics is a complex human concept rooted in cultural norms, values and social beliefs. Its integration into algorithmic processes must be based on a balance between different values and an understanding of various contexts. It is due to this balance that an articulate reflection on moral reasoning can be carried out. **Artificial intelligence systems (AIS)** make decisions based on algorithms and data which may be imperfect or biased. Ethics involves the power to do, the power to say, the power to tell and accountability (Ricœur, 2004). These capacities promote responsibility towards oneself, others and society. This dimension of ethics is very important, particularly within the technological revolution that is AI and we believe it is necessary to clarify it, both to measure its complexity and to reveal the contribution of ethics where choices have to be made.

“Ethics is a complex human concept rooted in cultural norms, values and social beliefs. Its integration into algorithmic processes must be based on a balance between different values and an understanding of various contexts.”

Background

The year 2022 marked a turning point in the world of AI. The refinement and deployment of major language models such as ChatGPT sent shockwaves through the AI field, and society as a whole. As soon as these models appeared, the scale and scope of their generative capacities became apparent, and many people came to consider the potential consequences of their large-scale deployment. Generative capacities of large language models such as GPT (OpenAI), Claude (Anthropic), PaLM (Google) or Llama (Meta) are attracting attention, leading to the frequent use of the term *generative AI*. These large, multimodal language models are now capable of performing, almost instantaneously, a wide range of complex tasks, from writing texts to generating images and writing computer codes. In this respect, ChatGPT was just the tip of the iceberg of generative AI (Peres & al., 2023). For instance, professional assessment tools used to test such AI models proved to be, in part, inadequate. This technological breakthrough not only changes the way we interact with computers, it also requires us to rethink the way we do things, ushering in a new era of human-machine interaction. The refinement of large-scale language models has given rise to countless debates, reflections and positions, within academia and industry, concerning the future of AI development. New models of generative AI increasingly simulate parts of the brain's functioning, offering a superior performance to what a human can do with regard to certain skills. Some authors even claim that they could achieve a form of «artificial morality» (Butlin & al., 2023).

All recognises that an important step has been taken, and that it has given rise to as many concerns as promises of improvement. A number of players in industry, and even in academia, have taken a joint position for greater responsibility regarding the risks of AI, considered by some as similar to extinction risks or on a par with pandemics and nuclear war (Center for AI Safety, 2023). This position can undoubtedly be described as catastrophist in nature, and it certainly revives the debate between inflationary and deflationary visions of AI (Maclure, 2020). Recently, the inflationary vision of AI has been widely supported by various players involved in the development of this technology. Some of the motivations behind this stance are reminiscent of the influence of movements such as **longtermism**, **transhumanism** and technological **singularitarianism**. Accumulating paradoxes, these currents contradict each other and, for the most part, claim to be based on science whereas, more often than not, they are the stuff of myths (Cassiani-Laurin, 2018). Behind these approaches, often nurtured and promoted by the technological ecosystem, lies a symbolism associated with catastrophic scenarios or a form of religiosity guaranteeing the promise of infinite virtual salvation. These currents strongly influence today's technological paradigm, linking up with scientific, technical and social ideals of our time.

“This technological breakthrough not only changes the way we interact with computers, it also requires us to rethink the way we do things, ushering in a new era of human-machine interaction.”

Social complexity and the production of ethics

The space technological developments occupy in our lives is disrupting human activities and influencing our social condition and environment, which in turn leads to social complexity. Because AI is shaping our present and future faster than any other invention beforehand, it is sometimes difficult to clearly see all its social effects. It is therefore proving complex to reflect deeply on these effects at the same tempo as AI develops. Changes depend on the way populations appropriate them, and require constant vigilance of how the use of technology evolves, takes root and is articulated within these very populations. It is already possible to observe how, in different societal cultures, **a new vision of the world and new relationships are taking root between humans and the biosphere.** These relationships are of a completely different nature to prior social relationships. In much the same way as the birth of bioethics followed advances in biology and medicine, we are now witnessing the emergence of a specific ethics applied to AI.

How do we define ethics, what should it consist of, and how can it be specific to the field of AI? Despite its great popularity in this field since 2016, in fact, ethics has been distorted, reduced and sometimes rendered powerless to play a leading role. At times misused, or instrumentalized, ethics has been the subject of numerous misunderstandings, and often hijacked for strategic and marketing purposes for the benefit of a productivist ideology. The report by the European Group on Ethics in Science and New Technologies entitled *Artificial Intelligence, Robotics and 'Autonomous' Systems* (EGE, 2018: 14) reminds us that, in the absence of a coordinated reflection on the ethical and social challenges of AI, we run a real risk of succumbing to a bazaar of principles and values shallowly justifying expected behaviours.

This lack of reflection is reminiscent of the first two waves of attempts to formalize the ethics of AI. Yet these steps were useful in defining principles that could frame AI systems (Georgieva & al., 2022) since a certain consensus emerged on which principles should be prioritised. Nonetheless, some have argued that this interest in ethics remains a sham, exposing industry strategies aimed at countering any kind of regulation. Others have highlighted a certain interpretative relativism referred to as ethics shopping, demonstrating a kind of incompatibility of standards that reduces the chances of comparison, competition and accountability (Floridi and Clément-Jones, 2019). These initial waves have contributed to the current predominance of an AI ethic that could be described as minimalist (Ménissier, 2023), the ambition mainly being limited to risk prevention and limitation. Faced with this observation, in the aim of placing ethics at the heart of AI, we need to reframe the production of diverse ethics and to identify the meaning of this production in the context of technological development, and the motives behind its constitution.

“At times misused, or instrumentalized, ethics has been the subject of numerous misunderstandings, and often hijacked for strategic and marketing purposes for the benefit of a productivist ideology.”

1

Ethics and AI ethics

*What kind of world do we want to live in?;
What values do we hold dear as a human community?;
How should we build this 'we'?*
(Dratwa, 2022)

1. Ethics and AI ethics

The popularity of ethics in various professional circles and among the general public lies in the perception that it is an easily accessible moral discourse, which can often be held intuitively and without the need for in-depth reflection. Ethics is thus recognized both as a philosophical discipline and as an activity that can be undertaken by anyone without any prerequisites, such as a background in philosophy. The paradox here is quite apparent and easy to dispel. Ethics is about doing good and avoiding evil: this kind of popular assertion is commonplace, and since everyone might have an idea of what is good, everyone might be able to do ethics simply by breathing. Yet, distinguishing good from evil is not at all obvious, and ethics cannot be reduced to just that. It is true that ethics contains the idea of appropriate behaviour, i.e. behaviour that may be good or in conformity with a value or rule (a principle). To know or determine the values or rules of good behavior, we call on the notions of moral ends and what constitutes good, but also of duty, obligation, decision, dilemma, virtue, etc. Yet, these notions are not intuitive.

1.1. Ethics versus checklists

Since 2017, particularly with the creation of the '23 Asilomar Principles on AI' and the Montreal Declaration for the Responsible Development of AI, AI ethics has taken the form of lists of principles and rules. In section 3.2, we will come back to the Montreal Declaration, the ambition being less a 'recipe' for ethical AI but more a call for a reflective and deliberative practice. However, this declaration, and all those that followed, were presented in the fixed, static form of a catalog of virtuous statements. All you have to do for an ethical AI is draw up the right list of principles (3, 5, 7, 9 or 10) and follow it to the letter, using an evaluation and control list. The form of a 10 Commandments Catalog, or checklist, is attractive because it is simple and makes one feel like they can access everything one needs in order to know how to act well, without having to commit to deep reflection. However, **ethics cannot fit in a list.**

“...ethics is an argued reflection that is subject to rational reasoning procedures.”

Ethics is first and foremost reasoned moral reflection. Since Antiquity, this way of thinking has been one of the fundamental branches of philosophy. The task of ethics is to determine the rules of life and action, to give recommendations and even injunctions for living well. Most people are accustomed to this way of doing ethics: it's what we find in popular formulas like *carpe diem* (literally, 'seize the day' or 'enjoy the present'). Other types of ethical rules also exist, such as moral duties: don't lie, don't do to others what you wouldn't want them to do to you, and so on. These rules are very common, finding their roots in both religious texts and popular culture. However, ethics is not about these rules (or any list of rules). What distinguishes ethics as a philosophical discipline, from a disciplinary field such as law, lies in its approach: ethics is an argued reflection that is subject to rational reasoning procedures. We enter the realm of ethics as a philosophical discipline when we question the rational foundations of a moral rule (whether religious or not), its logical consequences and its place in a coherent system of rules. From this point of view, drawing up lists of principles and norms is an ancillary, even minor, activity in ethics.

ETHICS AND LAW

Ethics and Law are sometimes confused, sometimes set in opposition. While possessing distinct normative spheres and mechanisms, they share a common language related to rules and norms of human conduct. Unlike law, ethics is incentive and refers to the identification and expression of values and principles which guide action. When shared, values and principles help to steer action and to provide the social justification required for its analysis. Its power resides in interpretation and allows actors to own the process of norm sense building (Verhaegen, 1984). Furthermore, ethics commit actors who are concerned by this norm sense building process in real life situations. On the other hand, law is about obligations and duties framed in rules and regulations. It stands on the power of rules to determine human action. The end goal of law is conformity of conduct to these legal rules and regulations, approved by a sole public external actor (a judge or a legislator) who punishes bad conduct. These two notions belong to different ranges, which must not be confused nor diminished by popular terms such as *soft law* and *hard law*, emptying them of their primary substance.

1.2. Ethical and technical standards

In philosophy, the concept of norm designates a statement that indicates what should be done; it can be used as a synonym for rule. In industry, and in AI in particular, the term norm is often used to designate technical standards. Whether it be an ethical norm or a technical norm, both types of norms refer to what we should do. But according to Aristotle, there is a crucial difference between ethical and technical norms: their relationship to what is produced.

In the case of ethical rules, action produces nothing external to the agent (the acting entity). The value of action lies not in a product but in the action itself: a generous action is good in and by itself, even if it has the effect of helping an ill-intentioned person. On the other hand, in the case of technical standards, the value of the action lies in the product itself. Thus, the action of designing a robust, stable, 'safe' AI system has no value in itself, it is the reality (the product) of the robust, stable or 'safe' AI system that has value. Designing reliable AI systems is not about ethics, it is simply about meeting a technical requirement that is the basis of engineering. Similarly, cybersecurity is not an ethical issue but a technical one. It is true that a computer scientist building an AI system in defiance of technical safety standards would be acting badly, contradicting ethical commitments to respect procedures, vigilance, honesty, etc. However, safety or reliability standards are technical, not ethical in and by themselves. Doing ethics requires reflection on what is morally valuable, on what gives meaning to our actions and our common life, on desirable or just ends, but also on what defines us as moral beings (agents). Thinking about the appropriate rules of action and behavior, philosophers have sought to justify them rationally, and have developed complex systems of principles and values. These are known as 'moral doctrines'.

“Designing reliable AI systems is not about ethics, it is simply about meeting a technical requirement that is the basis of engineering.”

MORAL DOCTRINES FOR THINKING ABOUT MORAL MACHINES

Traditionally, ethical (or moral) doctrines have been classified into three main families: virtue ethics, consequentialist ethics and deontological ethics. Virtue ethics (or aretic ethics¹) attributes the value of an action to the character of the agent: a person's generous action has value because it bears witness to that person's generous character, to her or his virtue. Consequentialist ethics determines the value of an action by assessing its consequences: the better the consequences of an action with regard to a chosen purpose, the better the action. For example, according to utilitarian thinking, an action is good if it maximizes pleasure or well-being. Finally, deontological ethics links the value of an action to respect for the rules, or the autonomy of individuals, even if it leads to bad consequences. For example, the action of not lying should not be motivated by the desire to maximize good consequences (and avoid bad ones), but solely by respect for the duty to tell the truth, or respect for others.

This way of presenting ethics is very popular in philosophy and beyond. In AI ethics, this ethical triad is sometimes used in practical cases, notably in the programming of ethical autonomous vehicles: should vehicles 'choose' the action that will increase the well-being (or minimize the harm) of people (consequentialist ethics) or the one that will respect their dignity (deontological ethics)? This question is thus raised when an agent is faced with a tragic choice, a dilemma where whatever she or he chooses, people's lives will be adversely affected. For example, in a situation where an accident is unavoidable and the vehicle can either continue on its way and kill X number of people, or veer into another lane and kill Y number of people, the question will then be whether X and Y, two numbers, are the only factor to be taken into account, or whether involving people in the accident while they were in another lane is morally acceptable. This is the famous 'trolley problem' (Foot, 1967/2002), popularized by MIT's Moral Machine Project, where the moral agent is not a person, but an autonomous vehicle (Awad, Dsouza, Kim, & al., 2018).

¹ The adjective aretic comes from the word *arête* (virtue in Greek).

2

The ethics of AI

That is why, when computer scientists legitimately ask how to apply principles, there is only one possible answer: we need to reflect on a given situation, and avoid automating moral reasoning.

2. The ethics of AI

The preceding paragraphs have clarified the notion of ethics and highlighted several possible applications of ethics to the field of AI. The distinction between ethical and technical norms is fundamental in defining the scope of ethics in AI. In addition, the outline of the three families of ethical doctrines has shown that an ethical problem in AI can be approached in several ways. But is AI ethics simply traditional ethics applied to AI?

2.1. Ethics applied to AI

In philosophy, ethics (1) and applied ethics (2) are distinguished in a way that seems very intuitive at first glance: applied ethics (2) is ethics (1) applied to a particular social practice (e.g. war), sector of activity (e.g. medicine), or object (e.g. the environment). Ethics is therefore general, and applied ethics is specific.

Ethics applied to AI would thus be a way of adapting the ethical doctrines of virtue (aretic), consequentialism or deontology to the problems raised by the development of AI². Two types of issues can then be addressed: the issue of good behavior by people who have a relationship with AI, whether in the research, development or deployment phases, and the issue of good behavior by machines, whether we lend them moral agentivity or not. However, another type of issue needs to be added to these two, one that lies at the frontier of the problem of AI governance: that of good behavior by institutions. In applied ethics, for each of these issues, we would have to apply principles previously elaborated within the framework of the great moral doctrines. Thus, in the case of the trolley problem, if we favor consequentialist ethics, we would say that the machine behaves ethically if it chooses to kill the fewest people; that the computer scientist has behaved ethically if he has programmed or trained the machine to kill the fewest people; and that the institution (e.g. the state) is good (has ethical concerns) if it has supervised the development and deployment of autonomous vehicles so that only consequentialist machines end up on the roads.

² See «The Ethics of the Ethics of AI » (Powers & Ganascia, 2020) on the challenges of ethics when applied to AI.

2.1.2 The origins of applied ethics: the case of bioethics

That said, this illustration of applied ethics is only true if we assume that ethical standards will be applied without modification. However, it is likely that they will be modified by contact with the object of reflection. For example, ethics applied to patients is not the same thing as general normative ethics: talking about respecting patients' dignity takes on its meaning in the specific context of a relationship with healthcare professionals. Let's take a simple case: should we tell a patient the truth about her or his state of health if there is no chance of them living for more than two months? A first possibility might be to apply the consequentialist principle on the basis that lying will increase well-being (Collins, 1927). This option, however, presents difficulties in calculating uncertain consequences. Another possibility might be to apply the deontological principle of respect for individual autonomy. However, the feeling that this choice could cause unnecessary harm would still persist. In such a situation, the ethicist has to take into account circumstances that are trivial yet crucial to finding an ethical solution: the truth has to be told, but it has to be told in a way that is caring and attentive to the patient's vulnerability (Higgs, 2007).

MEDICAL ETHICS AND BIOETHICS

Historically, medical ethics and bioethics were the first forms of applied ethics. They were born out of the trauma of Nazi experiments on deportees in concentration camps³. In the aftermath of the Second World War, the Nuremberg Doctors Trial, which ended in 1947, resulted in a judgment including the Nuremberg Code, which established 10 principles of ethical experimentation, among which informed consent. In 1964, the World Medical Association drew on the Nuremberg Code when it adopted the Declaration of Helsinki, which includes the principle of beneficence. From then on, medical ethics and bioethics became fully-fledged branches of normative ethics, based on four fundamental principles: beneficence, non-maleficence, autonomy and justice.

³ In the field of medicine, ethical concerns obviously go back further in time and are found in Hippocrates.

2.2. AI ethics versus “algorithmethics”

Two lessons dealing with the ethics of AI can be drawn from this historical review. Firstly, medical ethics is a synthesis of different moral doctrines. In the absence of consensus on a supreme value, it is necessary to balance the principles of virtue (aretic), consequentialism and deontology. Bioethics provides a model for applied ethics in general, and AI ethics in particular. Thus, the Montreal Declaration for the Responsible Development of AI (2018), one of the very first documents on the ethics of AI, follows in the wake of the Helsinki Declaration. In the Montreal Declaration, principles that had never been used before in the field of ethics are proposed for AI, such as the principle of human-machine solidarity and that of sustainable development, which takes into consideration the materiality of AI and its impact on the natural and social conditions of the existence of humans, animals, and all living things in general. This principle of sustainable development did not seem to be an issue in bioethics, but it becomes central to AI ethics.

Secondly, there is no such thing as ‘ready-made’ ethics: there are no formulae or algorithms for applying ethical principles. Some would like to ‘algorithmise ethics’ (doing ‘algorithmethics’) so that algorithms can reason ethically. However, the analysis of a moral situation (such as the trolley problem or the autonomous vehicle dilemma) is based on an effort of reflection. This effort of reflection enables us to formulate ethical principles, to define the essential characteristics of the situation, and then to apply the relevant principles according to a method of weighting or prioritising. **That is why, when computer scientists legitimately ask how to apply principles, there is only one possible answer: we need to reflect on a given situation, and avoid automating moral reasoning.** This is exactly where interdisciplinarity (or transdisciplinarity) is essential: it enriches the moral reflection of developers, and that of managers who take the decision to deploy artificial intelligence systems that have a social impact.

“...there is no such thing as ‘ready-made’ ethics: there are no formulae or algorithms for applying ethical principles.”

3

Methods of AI ethics

The role of the ethicist in a democracy is not to solve our most serious ethical problems, but to illuminate them so that democratic debate can take place in adequate terms that truly identify the crux (or cruxes) of the problem (Weinstock, 2006).

3. Methods of AI ethics

Ethics is a search for the moral meaning that guides existence and organizes social life. Ethical reflection is thus an effort to understand ourselves, to know ourselves as a collective, and to identify the values that stitch our lives together. We shouldn't think that values are so specific to one culture that they are incomprehensible to others. Rather, values are those points of reference shared by all communities, but which can be appreciated, organised and hierarchized in a way specific to each. The questions we ask ourselves as members of a moral community concern our identity as a community (who are we?), the kind of society we want for everyone, and the kind of solidarity we want to build. This questioning is fundamental and forms the starting point for all ethical reflection.

3.1. Deliberative ethics: the example of the Montreal Declaration

To determine what kind of society we want to live in, and what ethical principles we want AI to be based on, we need to resort to a form of reflection that is collective, reasonable and inclusive: the method of deliberation. Deliberation is a way of arriving at a decision (a practical judgment) through an exchange of rational and reasonable arguments. The exchange of arguments takes place in a conversation between individuals who live in the same social context, but do not have the same experience, the same point of view, the same conceptions of life and the good. Nevertheless, in this conversation, each person tries to understand the others and to produce the best understanding of the problem being discussed. In this deliberative process, participants strive to define the values that underpin their social lives. Deliberation requires a plurality of viewpoints to open up the field of possibilities and build the commonalities that will enable us to better define this new togetherness with technology. In a way, it represents an intervention in the potential as well as the actual transformation of society, involving a diversity of actors: multistakeholders, even omnistakeholders, including citizens and those who don't use AI.

“Deliberation requires a plurality of viewpoints to open up the field of possibilities and build the commonalities that will enable us to better define this new togetherness with technology.”

This method of deliberation was used to put together the Montreal Declaration for the Responsible Development of AI (2018). This inclusive deliberative process set a precedent for future work on AI ethics by developing the capacity of citizens to reflect on the issues it raises. As Daniel Weinstock reminds us: «**The role of the ethicist in a democracy is not to solve our most serious ethical problems, but to illuminate them so that democratic debate can take place in adequate terms that truly identify the crux (or cruxes) of the problem**» (2006). The deliberative process does not end the search for ethical principles and norms, but it does give it direction and meaning. Ethicists then extend the deliberation with their expertise, striving to formalize and make more coherent the proposals that emerge from deliberations. The democratic process in which this deliberation took place made it possible to debate ethical principles and demonstrated its relevance: «It is up to human and collective intelligence to define the goals of social life and, based on them, the orientations of the development of artificial intelligence so that it is socially responsible and morally acceptable» (Montreal Declaration, 2018). This democratisation, or modification, of the locus of power in technological decision-making takes greater account of the actions, needs and values of societies (Feenberg, 1999). The key is to create the right conditions for open, committed deliberation on technological issues.

Clearly, developing the capacity for technological deliberation, innovation and adaptation is vital for social progress, but it must be coupled with participatory mechanisms that foster a dynamic process of learning about technology. This implies the creation of consultative social spaces where communities can assess technological needs and options, as well as their impact. Deliberative conditions are important to empower individuals and communities to make meaningful choices about technology, to move away from being passive users or subjects of technology to becoming active agents, constructively shaping patterns of technological development. Such an approach offers citizens the means of emancipation, training and empowerment, rather than making them guinea pigs of technological experiments (Latour, 2001).

“This implies the creation of consultative social spaces where communities can assess technological needs and options, as well as their impact.”

Conclusion

As a set of digital techniques, AI is profoundly transforming certain social practices, forcing us to revisit ethical issues in the light of the societal values and priorities we wish to promote. The ethics of AI offer a unique opportunity to think together about the particular, the singular, the general and the irreducible, all the while building this togetherness. These are critical times. They offer us a unique opportunity to work towards establishing greater social justice, and to do so by exploiting to the fullest the possibilities of collective deliberation that takes into account scientific advances and the real needs of societies. This is the contribution that ethics can make to the need for societal harmonisation, so that everyone can live decently and benefit from new technologies.

“The ethics of AI offer a unique opportunity to think together about the particular, the singular, the general and the irreducible, all the while building this togetherness.”

Recommandations



Offer to actors in the science and technology sectors training activities on AI ethics that include a perspective from the humanities and social sciences.



Support the development of an ethical competency among AI researchers and developers, law makers, and other actors involved in the development, marketing, use and regulation of AI.



Through conferences, workshops and other activities opened to a large audience, **increase the awareness of citizens** with regard to the various social impacts and ethical challenges related to the development and deployment of AI.



Through a participative and interdisciplinary perspective that involves citizens as well as civil society organisations, **create spaces, instances and other mechanisms that promote reflection, discussion and evaluation of AI technologies and their use.**



Put in place in universities and public organisations interdisciplinary ethics committees specific to AI and computer or robotic technologies.

Lexicon

Generative artificial intelligence: “Generative AI (GenAI) is an Artificial Intelligence (AI) technology that automatically generates content in response to prompts written in natural language conversational interfaces. Rather than simply curating existing webpages, by drawing on existing content, GenAI actually produces new content. The content can appear in formats that comprise all symbolic representations of human thinking: texts written in natural language, images (including photographs to digital paintings and cartoons), videos, music and software code “ (UNESCO, 2023).

Artificial intelligence system (AIS): “a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. It uses machine and/or human-based inputs to perceive real and/or virtual environments; abstract such perceptions into models (in an automated manner e.g. with ML or manually); and use model inference to formulate options for information or action. AI systems are designed to operate with varying levels of autonomy. “ (OECD, 2019).

Longtermism: A view which holds that the positive influence of the long-term, or even very long-term, future is a major moral priority of our time, and that the prevention of existential risks to humanity should be prioritized, particularly in the allocation of resources (Boddington, 2023; MacAskill, 2022). Based on the premise that the vast majority of human beings are probably not yet born (MacAskill, 2022), and then relying on consequentialist reasoning, longtermists claim that the survival of civilization in the very long term, by enabling the existence of these several billion future lives, would ultimately increase tenfold the total value produced in the world (Boddington, 2023).

Transhumanism: “The intellectual and cultural movement that affirms the possibility and desirability of fundamentally improving the human condition through applied reason, especially by developing and making widely available technologies to eliminate aging and to greatly enhance human intellectual, physical, and psychological capacities.” (Bostrom, 2003).

Singularitarianism: “A movement related to transhumanism, postulating the advent of the singularity (Cassiani-Laurin, 2018), i.e. a moment in history when the accelerated evolution of technology, and AI in particular, would have such profound impacts that it would permanently transform human life, primarily through the fusion of human with machine” (Kurzweil 2005).

Bibliography

- Awad, E., Dsouza, S., Kim, R. & al. (2018). The Moral Machine experiment. *Nature*, 563(7729), 59–64. <https://doi.org/10.1038/s41586-018-0637-6>
- Boddington, P. (2023). Towards the Future with AI: Work and Superintelligence. In *AI Ethics. Artificial Intelligence: Foundations, Theory, and Algorithms*. Springer, Singapore. https://doi-org.acces.bibl.ulaval.ca/10.1007/978-981-19-9382-4_10
- Bostrom, N. (2003). The Transhumanist FAQ: A General Introduction. (Version 2.1). *World Transhumanist Association*. <https://nickbostrom.com/views/transhumanist.pdf>
- Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S.M., Frith, C., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M.A.K., Schwitzgebel, E., Simon, J. & VanRullen, R. (2023). Consciousness in Artificial Intelligence: Insights from the science of consciousness. *arXiv preprint arXiv:2308.08708*. <https://doi.org/10.48550/arXiv.2308.08708>
- Cassiani-Laurin, F. (2018). *La singularité transhumaniste comme mythe eschatologique contemporain : l'utopie technologique en tant qu'achèvement du projet moderne*. [mémoire de maîtrise, Université du Québec à Montréal]. Archipel. <https://archipel.uqam.ca/11714/>
- Center for AI Safety. (2023). *Statement on AI Risk*. <https://www.safe.ai/statement-on-ai-risk>
- Collins, J. (1927). Should doctors tell the truth?. *Harper's Monthly Magazine*, 155, 320–6.
- Montréal Declaration. (2018). The Montréal Declaration for a Responsible Development of Artificial Intelligence. <https://montrealdeclaration-responsibleai.com/the-declaration/>
- Dratwa, J. (2022). *Ethics of Transitions: What World Do We Want to Live in Together?*. ISTE.
- European Group on Ethics in Science and New Technologies (EGE). (2018). Statement on Artificial Intelligence, Robotics and “Autonomous” Systems. 24p. <https://www.unapcict.org/resources/ictd-infobank/statement-artificial-intelligence-robotics-and-autonomous-systems>
- Feenberg, A. (1999). *Questioning Technology*. Routledge
- Floridi, L. (2023). *L'éthique de l'intelligence artificielle : Principes, défis et opportunités*. Mimesis
- Floridi, L. & Clement-Jones, T. (2019). The five principles key to any ethical framework for AI. *Tech New Statesman*.
- Foot, P. (2002). The Problem of Abortion and the Doctrine of the Double Effect. In *Virtues and Vices: and other essays in moral philosophy*, Oxford University Press. <https://doi-org.acces.bibl.ulaval.ca/10.1093/0199252866.003.0002>. (Original work published in 1967).
- Georgieva, I., Lazo, C., Timan, T., & van Veenstra, A. F. (2022). From ai ethics principles to data science practice: a reflection and a gap analysis based on recent frameworks and practical experience. *AI and Ethics*, 2(4), 697–711. <https://doi.org/10.1007/s43681-021-00127-3>
- Higgs, R. (2007). Truth Telling. In R. Rhodes, L.P. Francis & A. Silvers (eds.), *The Blackwell Guide to Medical Ethics* (p.88–103). Blackwell Pub. DOI:10.1002/9780470690932
- Koniakou, V. (2023). From the “rush to ethics” to the “race for governance” in Artificial Intelligence. *Information Systems Frontiers*, 25(1), 71–102.
- Kurzweil, R. (2005). *The singularity is near : when humans transcend biology*. Viking.
- Latour, B. (2001). *L'espoir de pandore : pour une version réaliste de l'activité scientifique*. La Découverte et Syros.

- Latour, B. (2001). *L'espoir de pandore : pour une version réaliste de l'activité scientifique*. La Découverte et Syros.
- MacAskill, W. (2022). What is longtermism?. *BBC*. <https://www.bbc.com/future/article/20220805-what-is-longtermism-and-why-does-it-matter>
- Maclure, J. (2020). The new AI spring: a deflationary view. *AI & Society: Journal of Knowledge, Culture and Communication*, 35(3), 747–750. <https://doi.org/10.1007/s00146-019-00912-z>
- Ménissier, T. (2023). Les quatre éthiques de l'intelligence artificielle. *Revue d'anthropologie des connaissances* [En ligne], 17(2). <https://doi.org/10.4000/rac.29961>
- Nussbaum, M. C. (2012). *Philosophical interventions : reviews, 1986-2011*. Oxford University Press.
- OECD. (2019). *Artificial Intelligence in Society*. Éditions OECD, Paris, https://www.oecd-ilibrary.org/fr/science-and-technology/artificial-intelligence-in-society_eedfee77-en
- Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice. *International Journal of Research in Marketing*. <https://doi.org/10.1016/j.ijresmar.2023.03.001>
- Powers, T.M. & Ganascia, J.G. (2020). The Ethics of the Ethics of AI. In Dubber, M.D., Pasquale, F. & Das, S. (eds.), *The Oxford Handbook of Ethics of AI* (Oxford Academic, p.26-51). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190067397.013.2>
- Ricoeur, P. (2004). *Parcours de la reconnaissance : trois études* (Ser. Les essais). Stock.
- UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- UNESCO. (2023). Guidance for generative AI in education and research. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Verhaegen, J. (1984). Notions floues et droit pénal. In C. Perelman & al. (eds.), *Les notions à contenu variable en droit* (p. 7-19). Bruylant.
- Weinstock, D. (2006). *Profession, éthicien*. Presses de l'Université de Montréal.



obvia

www.obvia.ca