Including a Social Perspective In AI Ethics: The Contribution of a Dialogue Between American Pragmatism and Critical Theory

*Andréane Sabourin Laflamme, André-Laurendeau College (*Canada)

Frédérick Bruneault, Université du Québec à Montréal André-Laurendeau College (Canada)

International Conference on Computer Ethics: Philosophical Enquiry (CEPE) 2023, Chicago, IL

Keywords: AI ethics, moral philosophy, philosophy of technology, applied ethics, American pragmatism

Abstract

Throughout the history of moral philosophy, the theoretical postures have been privileged. Modern ethics is no exception and is indeed characterized by the predominance of voluntarist and universalist frameworks, which are primarily concerned with the actions of the moral agent, with no real regard for the conditions of possibility necessary for the effective realization of moral actions. Recent developments in applied ethics have shown that an integral application of classical ethical frameworks does not adequately address the new moral dilemmas emerging from our different spheres of activity. Artificial intelligence (AI) ethics once again demonstrates the inadequacy of traditional ethical frameworks to deal with the many ethical issues related to the pervasiveness of AI systems. Indeed, the dominant theories in ethics fail to take account of the shared responsibility that characterizes the moral obligations we have towards AI systems. The particularity of pragmatist ethics is that it aims at a practical intervention without however renouncing the conceptual clarifications necessary for such an intervention. We will demonstrate how the characteristics of pragmatist ethics avoids certain pitfalls in AI ethics and provides a conceptual framework particularly well suited to address the ethical issues related to the increasing use of AI systems in our societies.

Over the past few years, faced with the increasing integration of artificial intelligence systems (AIS) in our different spheres of activity, the radical transformations that the use of these technologies is causing in on our ways of functioning and the many ethical issues associated with these changes that have been identified (Tsamados et al., 2021), we have been witnessing the proliferation of initiatives in AI ethics. There are now more than 130 characters, declarations, and various types of guidelines and codes of conduct related to AI around the world[1]. Academic publications dealing with these issues have also multiplied at an accelerated pace over the last decade, and we can say that AI ethics now forms a branch of applied ethics, just like bioethics, business ethics or environmental ethics. All agree that the analysis of the ethical issues associated with the growing use of AI is pressing, given the serious risks that the use of this technology poses to the protection of our personal data and privacy, the risks of algorithmic bias and discrimination, the infringement of individual autonomy, and the possibilities of surveillance and control that AIS allow. In the last decade, these questions have increasingly mobilized the attention of philosophers and ethicists. Several approaches in ethics seek not only to provide a theoretical analysis of

[1] Council of Europe. Initiatives on AI. https://www.coe.int/fr/web/artificial-intelligence/national-initiatives.

the issues associated with the use of AI in our societies, but also to establish a framework from which to guide the design, development and deployment of these systems. However, many of these approaches do not include explicit reflection on the social, economic and political context in which ethical issues emerge in relation to the use of AI, thus reproducing the tendency, inherited from the classical ethical frameworks in moral philosophy, to remain focused on the moral agent, which results in what we call a social deficit. We argue that many proposals in AI ethics implicitly rely on an instrumentalist conception of technology, considered as axiologically neutral, thus eluding the fundamentally social nature of technology and presenting in this sense what we will call a *sociotechnical deficit*. This gap, in the specific context in which AIS evolve, which is characterized by historical socioeconomic inequalities, a strong asymmetry of power relations, the predominance of the free market logic and the of the ideology of surveillance capitalism (Zuboff, 2019), not only undermines the relevance of disembodied reflections on the moral dilemmas associated with AIS, but also hinders the practical purpose of certain ethical approaches such as those based on the adoption of codes of ethics founded on abstract principles. We argue that the understanding of ethics developed within American pragmatism has features that allow it to avoid the pitfalls of the classical voluntarist, universalist, and individualist frameworks in moral philosophy as well as the principlist approach in AI ethics, which is ubiquitous in the debate (as evidenced by the proliferation of statements of principles in AI ethics), and which also has significant deficits, among other things the challenges posed by the lack of organizational structures that would enable the implementation of these principles (Mittelstadt, 2019). We believe that the American pragmatist philosophy, as proposed by Charles Sanders Peirce, John Dewey, and by a contemporary authors such as Robert Brandom, since it defines ethics as an inherently social endeavor, avoids the dichotomies often presupposed and unquestioned by classical approaches in ethics — such as those implicitly established between technology and society — thus allowing to overcome the sociotechnical deficit that these classical approaches present. This also makes it possible to avoid the pitfalls related to the instrumentalist conception of technology[2] that prevails in the history of ethics and thus to rethink technology in its social context. In our view, by integrating into ethics a reflection on the systemic inequalities that characterize the context of technological development and an analysis of the structural and societal obstacles that can compromise the application of ethical principles in context, this type of posture also allows for the crossing between AI ethics and critical theory, discourses that often remain alien to each other. The pragmatist framework thus makes it possible to envisage AI ethics as a "situated ethics" (Rouvroy, 2022), open to multidisciplinarity and based on processes of democratic participation that allow for the legitimization of the practical solutions that may arise from social deliberative processes.

1. The socio-technical deficit of AI ethics

Since the middle of the 20th century, we have witnessed a practical turning point in moral philosophy which has given rise to the birth of applied ethics. In the post-war context, where minds are still freshly scarred by the exactions committed during the hostilities and where international legal frameworks are multiplying with the aim of preventing the reproduction of such horrors, several philosophers seek to mobilize ethics in order to analyze the moral dilemmas that emerge both in relation to political issues such

[2] As explained in Bruneault and Sabourin Laflamme (2021), the instrumentalist conception of technology stipulates that the development of technology is an enterprise instituted and controlled by humans who can therefore influence it as they see fit. This presupposes the thesis of technological neutrality. This position is based on the idea that technology is inherently neutral and that it is only indirectly morally charged, depending on what users decide to do with it, with the focus of value resting exclusively on that use.

as armed conflict, poverty, or social inequalities, and in specific professional contexts such as medicine, business, or law (van den Hoven, 2017: 12). Early initiatives in applied ethics, drawing on the conceptual legacy of moral philosophy that consolidated during modernity, initially attempted to mobilize theoretical frameworks such as deontologism and consequentialism in a deductive approach where moral theories are applied integrally to the specific issues under study. However, the limitations of such approaches soon become obvious. For Maesschalck (2010), a philosopher who has been interested in the historical transformations of ethics, the inadequacy of modern frameworks for the development of applied ethics stems, among other things, from the rationalist presuppositions that underlie them and that convey a universalist and voluntarist conception of ethics that hinders the analysis of moral dilemmas in their social context. On the one hand, the fact that it is not possible to settle once and for all the debate on which framework should be privileged calls into question the universalist postulate that classical ethical theories presuppose and compromises *de facto* the coherence of their application in context. On the other hand, still according to Maesschalck, the voluntarist character of modern ethics, whether deontological or consequentialist, reduces the field of moral reflection to the strict will to act, without regard to the conditions of possibility necessary for the effective realization of moral action or to the structural or social obstacles that could compromise this realization. For Hans Jonas, who takes issue with the presuppositions underlying modern ethics, this lack of interest in the historical processes and social contexts in which moral reflection takes place is a matter of what he calls *secular utopianism*, that is, a kind of faith or overconfidence in the power of circumstances (Maesschalck, 2010: 28). This moral hope betrays the rather individualistic and disembodied character of modern ethical frameworks, which are oriented towards the will that presides over the action of the moral agent, eclipsing from the field of ethics reflection on the specific social context in which moral action seeks to be embedded. Virtue ethics, which is considered the third classical approach in moral philosophy, and which constitutes a tradition stemming from Aristotelian philosophy, presents itself in a slightly different way, in the sense that it is not based the on will of the moral agent to favour the maximization of well-being, nor on his will to comply to universal moral duties, but rather in the development of moral qualities that allow individual fulfilment. Although virtue ethics posits the need for the development of *phronesis*, a form of wisdom or practical intelligence that develops with experience and that allows one to determine the action to be taken in a concrete situation, and although Aristotle specifies that individual fulfillment is intrinsically linked to the fulfillment of the State, virtue ethics is primarily focused on individual fulfillment and the reflection on the social context in which these virtues develop remains incidental. The history of moral philosophy is thus characterized by the omnipresence of theoretical frameworks that focus on the individual and individual moral decision-making, without any real regard for the specific social context in which ethical issues emerge. Modern ethics therefore presents what we will name a *social deficit*. Over the last century, many have condemned the disembodied character of ethics as it is traditionally taught in university philosophy faculties (Anscombe, 1958), namely as an intellectual exercise whose object of study is the logical coherence of the presuppositions underlying the theoretical frameworks under study. We can also think of Posner's (1998) virulent criticism of *academic moralism*, that he describes as a sterile posture fundamentally unfit to bring about any practical change whatsoever.

Faced with these criticisms, and in response to the inadequacies of classical ethical frameworks, several philosophers have proposed new approaches for the evaluation of contemporary ethical problems. For Maesschalck (2010), it is Hans Jonas (1979) who initiates, with his proposal of an ethics of life, the practical turn in ethics, which conceives moral responsibility in an empirical and concrete way, as a process that is inscribed in a particular socio-technical context and whose objective is the proactive

intervention in this precise context. According to Jonas, the fact that the technological environment in which we now evolve has profoundly transformed the relationship we have with our environment and with others, and has radically complexified the nature of the ethical issues we collectively face, sounds the death knell for the reign of modern rationalist individual and anthropocentric ethics. Indeed, not only must the concept of moral responsibility be broadened to encompass our obligations to the environment, but our responsibility must also be engaged with the consequences that our conduct may have on future generations. Approaches such as that of John Rawls, who in his *Theory of Justice* (1971) presents the political conditions necessary for justice as fairness, or that of Martha Nussbaum (2011), who in her capability approach insists on the need for social structures that allow for the functional realization of individual's potential, echo the new interventionist and social focus of contemporary ethics as well as the challenge to moral universalism that characterizes modern ethical approaches. Obviously, these proposals represent only a few examples of theories that are part of the practical turn in ethics, but the contribution of these approaches is that they allow to place moral issues in their specific social, political and economic context while helping to fill the social deficit that has marked the history of moral philosophy.

Given the ecosystem in which AI is deployed, which is characterized by the asymmetry of power between the producers of AIS and their users, and given the fact that algorithms have been shown to reproduce, accentuate and automate the systemic inequalities that exist in our societies, it is all the more essential that AI ethics avoids reproducing the social deficit that some approaches to ethics present. As we have shown (Bruneault and Sabourin Laflamme, 2021), although AI ethics is still a consolidating disciplinary field, it is possible to situate the different positions in the AI ethics debate on an argumentative spectrum that unfolds between two extremes that we have called *substantialism* and *instrumentalism*[3]. To situate the postures in relation to each other on this spectrum, we used the conceptual distinction drawn by Jocelyn Maclure (2018) between inflationary and deflationary perspectives in AI ethics. Inflationary perspectives are primarily concerned with the long-term consequences of AI development, including examining the risks and issues associated with the emergence of general AI (AGI), while the deflationary perspective, on the other hand, deals rather with the specific moral dilemmas currently posed by the different spheres of activity that are impacted by the use of weak AIS. We have argued that at the height of inflationism is the substantialist position. Essentially, substantialism postulates that technology is a substantial reality whose development is autonomous and obeys its own logic. In other words, substantialism is based on the thesis of technological determinism. In AI ethics, we can associate with substantialism postures such as those held by authors like Ray Kurzweil (2005), Nick Bostrom (2014), and Marx Tegmark (2017), who consider that given that the emergence of an AGI is a matter of necessity, the most pressing ethical problem is the *control problem* (Bostrom, 2014), which requires that we ask ourselves how to prevent AGI from acting in contradiction with human interests and values. These kinds of considerations received a lot of attention from the 2000s until about the middle of the 2010s, in the period immediately following the end of what we commonly call the AI winter, in what we propose to call the first wave of AI ethics. This type of posture, which can be described as "strong long termism", can be illustrated by the

[3] For reasons that we have already explained and that it is not possible to detail here, we believe that it is more useful to distinguish the different positions in AI ethics according to the postulates on which they are based in philosophy of technology than according to the usual distinctions in ethics that are based on the classical frameworks in moral philosophy.

publication of the open letter published in 2015 by the Future of Life Institute[4] and signed by, among others, Elon Musk, Nick Bostrom and physicist Stephan Hawking, all three of whom publicly warned of the existential risks associated with the emergence of a superintelligence. If these preoccupations have been overshadowed in recent years by more deflationary positions — that consider that the emergence of an AI that would pose an existential risk to humanity is rather unlikely and that it is currently the weak AIs that pose the most pressing ethical issues such as the infringements of privacy and autonomy (Maclure, 2018), the risks of bias and discrimination, and the possibilities of surveillance and control that AIS allow — inflationary discourses has dramatically resurfaced in 2023 along with the growing concerns around the multiplication of powerful generative AI models such as GPT-4, Google's Bard and Microsoft's Copilot. This renewed interest for inflationism in AI ethics has led to the new publication by The Future of Life Institute[5], urging us to pause the development of the training of AI systems more powerful than GPT-4, by evoking among other things the risk of loss of control of our civilization. Andrew Feenberg (2010), who distinguishes theories in philosophy of technology according to concepts similar to those we use, also associates some critical thinkers such as Jacques Ellul with substantialism. Although the rather technophile stance of authors such as Kurzweil, Bostrom, and Tegmark contrasts with the anti-utopian stance of philosophers such as Ellul (2012), the fact remains that in Feenberg's view, the proponents of transhumanism, technical progress is in some sense a necessity and occurs autonomously and independently of politics or society. Feenberg also associates Heidegger's and, to a lesser extent, Marcuse's approach with substantialism.

At the other end of the spectrum, at the height of deflationism, is instrumentalism, a posture that sees technology as a mere means, devoid of values and inherently neutral, the idea being that the value we can attribute to technology depends on the use made of it. It is therefore sufficient to frame technology in such a way as to limit the malicious uses we can make of it. In AI ethics, but also in the legal debate about AI, this type of approach is common. In this perspective, to address the ethical issues related to AI, we need to make sure, as it is the case with all human actions and sectors of activity, that AIS are deployed in accordance with human rights and in respect of democratic values. Obviously, in a context where AIS infringes on human rights on a regular basis (Access Now, 2018) and where States are struggling to ensure that the deployment of these technological devices complies with their international obligations in terms of fundamental rights, this posture seems relevant. Yet, in our view, the problem with radically deflationary postures at the extreme end of instrumentalism is that they risk reproducing the classic social deficit of ethics. Indeed, strong instrumentalist positions, which maintain that technologies, and in this case AIS, are neutral tools and that their impact depends solely on how they are used, neglect to take into account that technology develops in a particular social context and that it constitutes a vector of values transforming the way we interact with each other and with the world. Traditionally, ethics has adopted a rather instrumentalist conception of technology. In this sense, Feenberg (2010) states that the problem is that these approaches in applied ethics generally consider technology as an intangible and neutral given. According to Feenberg, ethics neglects to consider the fundamentally social dimension of technology, with the effect that it produces disembodied reflections that do not properly address the existing power relations and sociopolitical dynamics that can compromise the realization of ethical principles in practice.

[4] Future of Life Institute. (2015). Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter. https://futureoflife.org/open-letter/ai-open-letter/.
[5] The Future of Life Institute. (2023). Pause Giant AI Experiments: An Open Letter. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Approaches to AI ethics that rely on such an instrumentalist conception of technology risk reproducing the social deficit that characterizes classical ethics, but this time in the form of what we name a *socio-technical deficit*, namely the tendency to interpret technology independently of its social context. The principalist current in AI ethics, like the one that dominated the field of bioethics in its early days (Beauchamp and Childress, 1979), most definitely also presents this shortfall. Indeed, as discussed in introduction, the problem of applying the ethical principles enshrined in the multiple charters, declarations, codes and other guidelines that have been published recently remains unresolved (Morley, 2021) and the instrumentalization of ethics for the purpose of *ethics* washing is unfortunately now commonplace in the AI industry. It also appears that classical ethical frameworks, because of their individualistic, voluntaristic and universalist character, are not suitable to deal with the ethical issues of AI either. In fact, the integral application of these frameworks often leads to an anthropomorphization of AI and to an ethical reflection that transposes the moral agentivity of the individual to the artificial agent (or often to autonomous cars) and can lead to rather narrow qthical questions such as whether we should program a deontological, a consequentialist or virtuous robots (Gibert, 2020). Although these reflections are not without interest, in this case given the recent multiplication of conversational agents such as ChatGPT, GPT-4 or the conversational assistant associated with the search engine Bing (which has moreover made the headlines because of its particularly problematic behavior (Rose, 2023), the fact remains that most of the AIS with which we interact, even if they raise many ethical issues, do not have a human form and do not reproduce human interaction as such (think for example of the numerous recommendation algorithms that generate the content that is presented to us online or of facial recognition systems). In fact, it is very hard to avoid anthropomorphizing AI when using classical ethical frameworks, which again demonstrates their limits in this new disciplinary field of AI ethics. It appears that the postures we have described elsewhere as *ethics as usual* in AI ethics (Bruneault & Sabourin Laflamme, 2021) risk constituting a new iteration of the social deficit of ethics, which takes the form of a socio-technical deficit. We believe that the pragmatist framework is a particularly promising avenue for addressing this deficit since it not only provides an appropriate and systematic approach to the contextual analysis of AI ethics issues, but also allows us to consider ethics as a mode of individual action and social transformation. We will now present the epistemological foundations of pragmatism and we will then to demonstrate how, based on these principles, we can design a methodology that allows us to establish a process of analysis in applied ethics that avoids the classical social deficit of ethics and, more specifically, the sociotechnical deficit in AI ethics.

2. The contribution of pragmatism to AI ethics

Emphasizing pragmatically focused interpretation of language (rather than syntax or semantics), pragmatism draws on semiotics, the theory of signs, proposed by Charles Sanders Peirce (De Waal, 2013), a theory whose explanation goes far beyond the focus of this text. Moreover, as several scholars of the pragmatist tradition in philosophy point out (Talisse & Aikin, 2008; Misak, 2013), there is no single interpretation of the theoretical foundations of this school of thought. The purpose here, therefore, is not to account for the complexity of the debates surrounding these issues, but rather to demonstrate how an approach to AI ethics inspired by such a philosophical perspective overcomes the shortcomings of the

classical ethical frameworks identified above. A general presentation of the important features of the pragmatist approach will therefore be sufficient to achieve this goal.

One possible starting point for presenting pragmatism is the pragmatic maxim, as proposed by Peirce, that the meaning of a concept or idea can be understood in terms of its practical effects on our experience (Misak, 2013). The meaning of a concept or idea is determined by its practical consequences and its use in our lives. This implies that the meaning of a concept or statement is not fixed or absolute, but rather is determined by its use in concrete situations. By focusing on practical consequences, we can avoid getting bogged down in abstract or metaphysical debates, and focus on how to use ideas to improve our lives and solve real-world problems (De Waal, 2013). Peirce (1877) explores in his text "The Fixation of Belief" how individuals form and maintain their beliefs. According to Peirce, belief is a central element of human life. We form beliefs in order to make sense of the world around us and guide our actions (Bacon, 2012). However, Peirce argues that belief is also inherently unstable and subject to doubt and uncertainty. In order to overcome this instability, individuals engage in a process of inquiry that aims to fix their beliefs in a more secure and reliable manner. Peirce identifies four methods that individuals use to fix their beliefs: a- the method of tenacity, b- the method of authority, c- the method of a priori reasoning and d- the scientific method. The method of tenacity (a) is the most elementary and primitive way of fixing beliefs. It consists in clinging to beliefs simply because they seem right or because they have always existed. This method is prone to error and can lead to dogmatism and closed-mindedness. The method of authority (b) consists of fixing beliefs on the basis of the testimony of others who are considered experts or authorities. This method is also error-prone and can lead to blind faith and uncritical thinking. The method of a priori reasoning (c) involves fixing beliefs based on deductive reasoning from first principles or self-evident truths. This method can be useful in some contexts, such as in mathematics or logic, but it can lead to circular reasoning and over-reliance on first principles or deductions. The scientific method (d) involves formulating hypotheses, testing them through observation and experimentation, and revising them according to new evidence. By subjecting beliefs to rigorous testing and evaluation, the scientific method enables individuals to arrive at more accurate and reliable beliefs that are more likely to stand the test of time. For Peirce, individuals should strive to use the scientific method to arrive at the most accurate and reliable beliefs.

To fully appreciate the shift in perspective proposed by pragmatism, it is helpful to understand the influence of two scientific revolutions contemporary to the work of early pragmatists, namely the development of evolutionary theory and the use of statistical methods (Brandom, 2011). First, evolutionary theory provided pragmatists with a robust conceptual framework for thinking about the evolution of human knowledge in history. Before Darwin, most philosophers believed in a static, unchanging world governed by fixed laws and principles. However, Darwin's theory of evolution challenged this perspective by suggesting that the natural world was constantly changing and evolving in response to new environmental pressures. Darwin's theory of evolution proposed a naturalistic explanation for the diversity of life, challenging traditional religious and philosophical views of a fixed, unchanging world. This has had a profound effect on the way pragmatists have approached questions about knowledge, truth, and reality. This perspective suggests that knowledge and truth are not fixed and absolute, but constantly evolving and changing over time. Evolution thus appears as a continuous process of adaptation and adjustment to new environments and changing circumstances, a view that corresponds to the importance pragmatists place on practical problem solving and experimentation. Darwin's theory of

natural selection emphasized the importance of experience in shaping the development of living organisms. This emphasis on empirical evidence and experimentation, coupled with the rejection of claims to absolute truth, has become a central tenet of pragmatism. Darwin's theory of evolution provided pragmatists with a model for understanding how knowledge evolves in ways similar to living species. Just as species adapt and change over time in response to their environment, our ideas and beliefs adapt and change in response to new experiences and information. Dewey was deeply influenced by Darwin's theory of evolution, and he considers the concept of adaptation central to the structure of his philosophy (Hildebrand, 2008). Dewey sees all knowledge as inherently practical and the relevance of a theory is measured by its ability to solve problems that arise in lived experience. For Dewey, knowledge is not a fixed set of principles, but rather a process of inquiry and experimentation that is constantly evolving in response to new problems that arise. Darwin's ideas on evolution and adaptation have also had a significant impact on the ethics and social theory of pragmatists, emphasizing the importance of social experimentation and adaptation to changing conditions (Stuhr, 1998). They see morality and social norms as evolving, rather than as fixed and absolute. This is embodied in Dewey's idea of democracy as a form of collective life based on social experimentation, rather than simply a type of political regime (Hildebrand, 2008).

Second, the rise of statistical reasoning also played an important role in the development of pragmatism, as it provided a new way of understanding the world consistent with the recognition of uncertainty and contingency that permeates human experience and the world itself (Brandom, 2011). Pragmatists view the world as inherently complex and our knowledge of it as always provisional and subject to revision. In this context, the rise of statistical reasoning has transformed the way we think about uncertainty. Statistics allows us to deal with uncertain and variable data by using probability theory to make inferences and predictions. Pragmatists have seen statistical reasoning as a powerful tool for dealing with uncertainty by making sense of data, identifying patterns and trends, and testing hypotheses. The rise of statistical reasoning has contributed to the development of pragmatism because it offers a way to deal with uncertainty by drawing inferences from random phenomena.

All of this leads to the main features of the theoretical underpinnings of pragmatism, the first of which is fallibilism (Talisse & Aikin, 2008), the idea that all human beliefs and knowledge claims are inherently fallible and subject to revision. This means that no belief is absolutely certain or indubitable and that all beliefs are subject to correction or improvement based on new evidence or experience. This point of view is opposed to more traditional epistemological postures, such as foundationalism[6] and coherentism[7] , according to which knowledge is constructed on the basis of fundamental and indubitable beliefs that can neither be questioned nor revised. Thus, fallibilism rests on the recognition that human beings are fallible and that all of our claims to knowledge are based on imperfect and limited information. Fallibilism has important implications on how we think about the nature and scope of knowledge. It suggests that knowledge is always open-ended and that there are no absolute or final truths. Fallibilism also has important implications on how we think about the role of doubt and uncertainty in our lives. Rather than being seen as obstacles to knowledge, doubt and uncertainty are seen as tools to guide inquiry and

[6] Epistemological theory according to which knowledge is based on incontestable and obvious foundations that serve as a starting point to justify our beliefs.
[7] Epistemological theory according to which knowledge is based on a coherent and logical system of beliefs, without the need for it to be based on incontestable and obvious foundations.

discovery, helping us to identify areas in which our beliefs need revision, in order to avoid dogmatism and narrow-mindedness.

This brings us to a discussion of the conception of truth underlying fallibilism. For Peirce, truth is the view that would be held by the participants in an inquiry process if it were to continue indefinitely - or at least until there was no longer any reason to believe that further inquiry could provide further evidence that might change that view. In a similar sense, John Dewey argues that truth is a process of inquiry and verification, rather than a fixed or absolute concept, though Dewey prefers to speak of "*warranted assertion*" rather than truth (Brandom, 2011). For Dewey, we use our "warranted assertions" to solve problems and make sense of the world, rather than as a final or ultimate goal. For pragmatists, truth (or "justified claims") is not an abstract or transcendent concept, but emerges from our practical experience and interaction with the world. Truth does not exist in isolation, but rather is a product of the ongoing process of inquiry and experimentation in which we engage as we try to make sense of our experiences. For his part, Dewey emphasizes the idea that our beliefs are not merely passive representations of reality, but rather active tools we use to interact with the world and achieve our goals (Brandom, 2011).

This is why one of the key concepts in Dewey's philosophy is the idea of experience. For Dewey, thought emerges from our practical experience of the world, rather than from abstract or theoretical reasoning. He argues that we cannot fully understand the world or solve practical problems through purely abstract or theoretical thinking, but rather must rely on our practical experience and interaction with the world. Dewey sees experience as the foundation of human knowledge and understanding. For him, experience is not simply a collection of sensory impressions or isolated events, but a dynamic and ongoing process that involves interaction between an organism and its environment (Hildebrand, 2008). Experience, for Dewey, is a continuous process that involves both the individual and the world around them. Moreover, experience is not limited to the individual, but is a social and cultural phenomenon (Brandom, 2011). Dewey argues that our experiences are shaped by our interactions with others and the social world to which the individual belongs. For him, the development of language, culture, and social institutions is part of the ongoing process of experience, and these elements help shape the way we understand the world (Campbell, 1998). Another of the key concepts in the pragmatists' conception of knowledge is the idea of inquiry (Misak, 2013). Inquiry is the process of actively seeking new knowledge through experimentation, observation, and reflection (Hickman, 1998). In this process, we develop theories and hypotheses about the world, which we then test through practical experimentation and observation. Through this process of inquiry, we can gradually refine and improve our understanding of the world and develop more effective strategies for achieving our goals. According to pragmatists, inquiry is not a purely cognitive process, but rather a practical, embodied process. This process is not linear or predetermined, but is open and responsive to changing circumstances. As knowledge is always situated in specific social, cultural and historical contexts, pragmatists emphasize the fundamentally social dimension of any form of research (Talisse & Aikin, 2008). It is important to understand that knowledge is not only the product of individual cognition, but rather the result of a collective and social process that involves interaction and dialogue with others. It is particularly interesting to underline here this social component of inquiry, since individuals cannot acquire knowledge only by themselves, it is always developed through a process of social interaction.

The interest of the pragmatist approach in ethics (Bruneault & Sabourin Laflamme, 2022) revolves around three main characteristics of the pragmatist approach identified by Keulartz et al. (2002), namely 1- anti-foundationalism, 2- anti-dualism and 3- anti-scepticism. 1- Anti-foundationalism refers to the fallibilism of pragmatist approaches. In ethics, it is, therefore, advisable to avoid making one of the major principles (defined by the classical approaches in ethics) the absolute principle which would then be at the basis of all our moral obligations. It seems preferable to mobilize these different theoretical frameworks to feed the reflection on specific ethical issues (Pappas, 1998), especially those raised by the development of AI. 2- Anti-dualism leads pragmatists to reject the main oppositions that characterize the main philosophical debates, such as those between mind and matter, facts and values, knowledge and appearances, or the individual and society. These distinctions may be useful for analyzing specific situations and for thinking about the problems generated by the development of AI, but they do not oppose substantial realities that are independent of each other. Rather, these terms are constantly in a co-constitutional relationship and do not exist independently of this co-constitution. This is particularly important, in our view, for thinking about the specific problems raised by AI, especially with regard to the distinction between individual and society, but also that between technology and society (Stuhr, 1998). 3- Anti-skepticism stipulates that the questioning of our knowledge must be motivated by specific problematic situations and that it is counter-productive (if not impossible) to start from a position of radical and absolute doubt. All of our knowledge can be questioned, but not all of it can be questioned at the same time. Pragmatist ethics, therefore, rejects the idea of a *tabula rasa* of ethical principles, especially when it comes to thinking about new technologies such as AI and the moral dilemmas that it can generate. Ethics, like all forms of knowledge, necessarily relies on inquiry (Pappas, 2008). In the ethical assessment of AI, inquiry takes the form of deliberation around specific problems raised by these technological innovations, always from a background of social and conceptual practices.

This is why we think that the pragmatist approach to AI ethics is particularly interesting, because it allows us to respond to the inadequacies of traditional approaches to ethics, inadequacies that become preponderant when we examine the problems raised by AI, as we have seen above in section 1. More specifically, the pragmatist conception allows us to remedy the individualistic tendency of these traditional ethical approaches by making, as we have just emphasized, a central place for the social dimension of the knowledge processes. It is moreover this reintroduction of the social question that allows us, in our opinion, to note a parallel between this pragmatist conception of ethics and an approach inspired by critical theory such as that of Feenberg, a parallel that we are now able to appreciate at its full value.

3. The contribution of a dialogue between pragmatism and critical thinking in AI ethics

The main features of the pragmatist approach to ethics lead us to argue that there are points of intersection between this approach and the one based on the critical theory tradition, despite the notable differences between them and the criticisms that some critical thinkers have made of pragmatism (Hildreth, 2009). In this sense, we fully endorse the thesis proposed by Larry Hickman (2007: 90) who states in his recent work that the theses supported by Feenberg are close to those of the pragmatist position in technology ethics because he:

“ a) move from an essentialist to a functionalist understanding of technology, b) developed a vigorous form of social constructivism, c) rejected a Heideggerian romanticism in favor of a naturalized technology, d) rejected the Critical Theorists’ notion of technology as ideology, e) accepted the idea that the project of Enlightenment rationality is not as much of a threat as the Critical Theorists had

imagined, f) proposed the idea that technical decisions are made within a network of competing factors in which one weighs various desired ends against one another, g) warned against the reification of the results of inquiry as if they had existed prior to inquiry (Dewey's ''philosophic fallacy''), and h) recast technology in a way that crosses the line between artifacts and social relations."

This is what leads Hickman to argue that the positions defended by the proponents of pragmatism and those proposed by critical theorists with regard to technological development - as far apart as they may have been in the past, notably because of the different epistemological postulates of these approaches - appear today to be much closer to one another. To take up briefly the elements identified by Hickman, it appears indeed that despite his deep affinities with the first thinkers of critical theory, first and foremost with Marcuse, Feenberg detaches himself from the technological substantialism that characterized their positions on the question, but also the ones of Heidegger and Ellul, which can be associated with an essentialist and therefore deterministic interpretation of technology. Feenberg argues that it is now necessary to support an anti-essentialist interpretation of technology (Feenberg, 1999: viii) - given more recent work in the philosophy of technology (Winner, 1980; Verbeek, 2005). In doing so, he seeks to extend the social constructivism already championed by early critical theorists to the question of technology and its co-constitution with society (Hickman, 2007). Such a development in Feenberg's thinking most certainly echoes the pragmatist position described in section 2 above. This critique of essentialism and technological determinism is also reflected in the reinsertion of technology into the broader history of social evolution. For Feenberg, technology is presented not as a simple instrument of control over nature, but as a social enterprise in its own right that cannot be reduced to a simple tool of ideological domination, nor to a simple incarnation of a purely instrumental reason in the service of the interests of the dominant class. Feenberg seeks to reintegrate technology into the lived world, rejecting a monolithic interpretation (Hickman, 2007) in favor of a finer analysis of the various, often conflicting, interests that preside over its development, the underlying motives and its insertion into human practices. This leads Feenberg closer to the interpretation of technology defended by pragmatists. Finally, Feenberg appears much closer to the pragmatists than to the classical critical theorists when he questions the tendency to reify the opposite terms of the great dichotomies, notably the one that opposes technology and society, which translates into a more nuanced interpretation of the social and cultural dimension of technological objects, which ultimately rests on challenging the supposedly substantial nature of the opposition between facts and values. Such an analysis is greatly influenced by the thinking of the last generations of philosophers of technology who insist on the idea of values are *embedded* in technological objects (Ihde, 1990; Verbeek, 2005). In short, despite the historical divergences of the philosophical traditions from which they originate and despite the differences that still oppose them, both on the epistemological and political levels, we believe that the pragmatist approaches to technology and those stemming from critical theory (Margolis, 1998), as defended by Feenberg, are now sufficiently close in several respects to hope to think of a fruitful theoretical dialogue between them, but above all, of a practical complementarity in the ethical evaluation of the issues raised by AI.

We believe that adopting a pragmatist framework in AI ethics allows for the bridging of ethics and critical theory, a passage that proves more difficult when we adopt classical frameworks in ethics that tend to elude the social nature of technology and to eclipse from the analysis of moral dilemmas the social, political and economic context in which they are embedded. The integration of considerations about the new forms of oppression and domination that AI can provoke in our societies within the very heart of ethical reflection, traditionally circumscribed to reflection on the principles that should guide individual

action, allows us to analyze the obstacles to the realization of moral actions in context and to consider possible solutions to circumvent them. Pragmatism also makes it possible to consider ethics as an ongoing, collective and contextual deliberative exercise and not as a one-off, individual and disembodied reflection. This posture also has the advantage of allowing us to think of moral responsibility as a shared duty, which is all the more appropriate for ethical reflection that concerns AI, given that AI is designed, developed and deployed by a multiplicity of actors among whom moral obligations are distributed (Floridi, 2013: 261). We believe that the pragmatist approach is a particularly promising avenue for moving from principles to practices in AI ethics (Morley, 2020) and for developing legitimate, democratic, and transparent processes for conducting ethical assessment of specific technological devices that identify the ethical risks involved in these systems and possible avenues for mitigating them. Adopting a pragmatist framework allows us to think of ethics as a deliberative and democratic process that requires the participation of all stakeholders, including marginalized groups and minorities, in a collective exercise where all voices must be heard. Viewed in this way, AI ethics can aspire to be embedded in practice as a mode of social intervention and as a source of legitimate normativity in AI governance. The pragmatist approach, coupled with a critical theory of technology, as proposed by Feenberg, allows us to envisage how ethics can contribute, by proposing a collective normative reflection on the values that should guide technological development, and to re-politicize technology. Indeed, such a posture opens the way for ethics to contribute to what Feenberg characterizes as "technical democracy (Feenberg, 2010). This type of stance echoes some approaches in technology ethics such as *Value Sensitive Design* (Friedman & Hendry, 2019; Lucivero, 2016), an approach that relies on a collective thought process that seeks to integrate moral values into the design of technologies, and is explicitly in line with the pragmatist conception of ethics.

References

Access Now (2018). Human Rights in the Age of Artificial Intelligence. 1-40. https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf

Elizabeth, G. M. Anscombe (1958). Modern Moral Philosophy. *Philosophy*. 33(124). 1-19.

Beauchamp, Tom and James Childress (1979). *Principles of Biomedical Ethics*. Oxford University Press.

Bacon, Michael (2012). *Pragmatism. An Introduction*. Cambridge: Polity.

Bostrom, Nick (2014). *Superintelligence, Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Brandom, Robert B. (2011). *Perspectives on Pragmatism. Classical, Recent, & Contemporary*. Harvard University Press.

Bruneault, Frédérick and Sabourin Laflamme, Andréane. (2021). AI ethics debate: How can information ethics provide a framework which allows us to avoid usual conceptual pitfalls? An overview ». AI & Society Journal. 1-26. https://doi.org/10.1007/s00146-020-01077-w.

Campbell, James (1998). Dewey's Conception of Community. *Reading Dewey*. Indiana University Press.

De Waal, Cornelis (2013). *Introducing Pragmatism. A Tool for Rethinking Philosophy*. New York: Routledge.

Ellul, Jacques. (2012). *Le système technicien*. Paris : Cherche Midi.

Feenberg, Andrew. (2010). *Between Reason and Experience*. Cambridge: Massachusetts Institute of Technology.

Feenberg, Andrew. (1999). *Questioning Technology*. London: Routledge.

Floridi, Luciano (2013). *The Ethics of Information*. Oxford University Press.

Friedman, Batya and David G. Hendry.(2019). *Value Sensitive Design. Shaping Technology with Moral Imagination*. Cambridge: MIT Press.

Gibert, Martin., (2020). *Faire la morale aux robots. Une introduction à l'éthique des algorithmes*. Montréal : Atelier 10.

Hickman, Larry A. (2007). *Pragmatism as Post-postmodernism. Lessons from John Dewey*. Fordham University Press.

Hickman, Larry A. (1998). Dewey's Theory of Inquiry. *Reading Dewey*. Indiana University Press.

Hildebrand, David (2008). *Dewey*. Oxford: One World.

Hildreth, Roudy W. (2009). Reconstructing Dewey on Power. *Political Theory*. 37(6). https://doi.org/10.1177/0090591709345454.

Ihde, Don (1990). *Technology and the Lifeworld*. Indiana University Press.

Jonas, Hans (1979). *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation.* Suhrkamp.

Keulartz, Jozef et al (2002). Ethics in a technological culture. A proposal for a pragmatist approach. In Pragmatist Ethics for a Technological Culture. Kluwer.

Kurzweil, Ray (2005). *The Singularity is Near*. New York: Viking.

Lucivero, Federica (2016). *Ethical Assessments of Emerging Technologies*. New York: Springer.

Maclure, Jocelyn (2019). The New AI Spring: A Deflationary View. *AI & Society. A Journal of Knowledge, Culture and Communication*. https://doi.org/10.1007/s00146-019-00912-z.

Maclure, Jocelyn. et Saint-Pierre, M.-N. (2018). Le nouvel âge de l'intelligence artificielle : Une synthèse des enjeux éthiques. *Les cahiers de propriété intellectuelle*. 30(3). 741-765.

Maesschalck, Marc. (2010). *Transformations de l'éthique. De la phénoménologie radicale au pragmatisme social*. Bruxelles : P.I.E Peter Lang.

Margolis, Joseph (1998). Dewey in Dialogue with Continental Philosophy. *Reading Dewey*. Indiana University Press.

Misak, Cheryl (2013). *The American Pragmatists*. Oxford University Press.

Morley, Jessica et al. (2021). Operationalising AI ethics: barriers, enablers and next steps. *AI & Society*. https://doi.org/10.1007/s00146-021-01308-8.

Morley, J. *et al.* (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Sci Eng Ethics*. 26, 2141-2168. https://doi.org/10.1007/s11948-019-00165-5.

Mittelstadt, Brent (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1, 501-507.

Nussbaum, Martha (2011). Creating Capabilities: The Human Development Approach. Cambridge, MA: Belknap Press of Harvard University Press.

Posner, Richard. A. (1998). The problematics of moral and legal theory. Harvard Law Review, 111, 1637-1717.

Pappas, Gregory F. (2008). *John Dewey's Ethics. Democracy as Experience*. Indiana University Press.

Peirce, Charles Sanders (1877). The Fixation of Belief. *Popular Science Monthly*. 12(1). 1-15.

Rawls, John (1971). *A Theory of Justice*. Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Rouvroy, A. et Manuel Zacklad (2022). L'éthique située de l'IA et ses controverses, *Revue française des sciences de l'information et de la communication*. 25. https://doi.org/10.4000/rfsic.13204.

Rose, Kevin. (February 2023). *A Conversation With Bing's Chatbot Left Me Deeply Unsettled*. New York Times. https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html.

Stuhr, John J. (1998). Dewey's Social and Political Philosophy. *Reading Dewey*. Indiana University Press.

Talisse, Robert B. & Scott F. Aikin. (2008). *Pragmatism*. Continuum.

Tegmark, Mark (2017). *Life 3.0, Being Human in the Age of Artificial Intelligence*. New York, NY: Vintage Books.

Tsamados, Andreas et al. (2021). The ethics of algorithms: key problems and solutions. *AI & Society*, 37, 215-230. https://doi.org/10.1007/s00146-021-01154-8.

van den Hoven, Jurgen, Seumas Miller, and Tomas Pogge (2017). *Designing in Ethics*. Cambridge: Cambridge University Press.

Verbeek, Peter-Paul (2005). *What Things Do. Philosophical Reflections on Technology, Agency, and Design*. The Pennsylvania State University Press.

Winner, Langdon (1980). Do Artifacts Have Politics? *Deadalus*. 109(1). 121-136.

Zuboff, Shoshana.(2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books Ltd.