How to Teach AI Ethics?

A Pragmatist Approach to Ethical Competency

Andréane Sabourin Laflamme and Frédérick Bruneault

Source: Wacomka/iStock

Artificial intelligence systems (AIS) are increasingly becoming an integral part of our various spheres of activity and include things like recommendation algorithms that generate the content presented in our news feed on social networks, voice assistants, autonomous cars, methods for screening certain cancers, conversational robots used for tutoring purposes, or the identification of students at risk of failure or dropout. Although this technology is promising in several respects, the numerous research studies that have been devoted to the issue over the past decade (Tsamados, 2021; Bruneault and Sabourin Laflamme, 2021) have identified several ethical and social issues associated with its use, such as the risks of bias, algorithmic discrimination or violation of privacy rights. These issues concern both the people who design the algorithms, the companies that use them, the users, and the people who work in the sectors transformed by the use of AIS. Given the automation of professional practices that is expected in the next few years (Conseil interprofessionnel du Québec, 2021), most people who graduate from a Quebec college will eventually be confronted with these issues, whether in their work, their academic career, their civic activities or their personal life.

As early as 2018, in its Digital Action Plan (DAP), the ministère de l'Éducation et de l'Enseignement supérieur (MEES) stipulated that digital competency, as an essential component of digital citizenship education, must necessarily include digital and artificial intelligence (AI) ethics education. In 2019, the Digital Competency Framework, which is a continuation of the DAP, reiterated this position by defining the first dimension of digital competency—targeted as a transversal competency—by the ability to act as an ethical citizen in the digital age. In addition, Dimension 11 of digital competency specifically addresses the ability to think critically about digital technology. Ethical and digital competencies are also an integral part of the new Référentiel québécois des compétences du futur [Quebec Competencies of the Future Framework, Ed.] (2022). Although the question of ethical and social issues related to AIS generates a lot of interest in the academic and institutional communities as well as in the technology development industry, our research has allowed us to conclude that the training offer in AI ethics is currently relatively limited, both in colleges and universities, and that there is no consensus on the characteristics that such training should have. The project to design an AI ethics competency framework[1] in higher education (Bruneault and Sabourin Laflamme, 2022), funded by the Pôle montréalais d'enseignement supérieur en intelligence artificielle [Montreal hub for higher education in artificial intelligence, Ed.] (PIA), is in line with these imperatives and aims to fill these gaps. Indeed, in the current context characterized by the ubiquity of AIS and the multiplication of ethical issues related to their deployment, it appears essential to identify the components of an adequate and complete higher education training in AI ethics. The objective of the framework is to provide a model for the development of different types of training in AI ethics and to ensure that they are integrated into the initial curriculum of the different study programs, in complementary courses or in extracurricular activities. The framework has been designed to be adaptable to many teaching contexts depending on the objectives, level and nature of the training in which these components are to be integrated.

Some methodological points of reference

The framework was produced after a three-part process. First, a review of the literature in AI ethics, and more specifically in the teaching of AI, digital and technology ethics, was carried out. We were interested in a more general way in the different expressions of the definition of ethical competency in the scientific literature. In parallel, a review of the course offer in AI ethics in Quebec colleges and universities was conducted in order to provide a general portrait of education in AI ethics and to measure the diversity of pedagogical approaches deployed in this context. Finally, 26 individual interviews were conducted with people who have expertise in teaching ethical and social issues related to AIS. Ethicists, sociologists, lawyers, and computer scientists were interviewed regarding their teaching methods, and their opinions on the most promising approaches for adequate training in AI ethics in higher education were collected. In addition, in collaboration with our partners at Eductive, we conducted a living lab, i.e., a co-construction exercise that brought together people from various backgrounds (research, college and university teaching, pedagogy, technology industry, etc.). We also facilitated a workshop with members of the *Ethics, Governance and Democracy* research group of the International Observatory on the Societal Impacts of AI and Digital Technology (OBVIA), during which we presented a preliminary version of the competency framework and gathered the opinions, comments and suggestions of specialists in research and higher education on the relevance of the model proposed in the document.

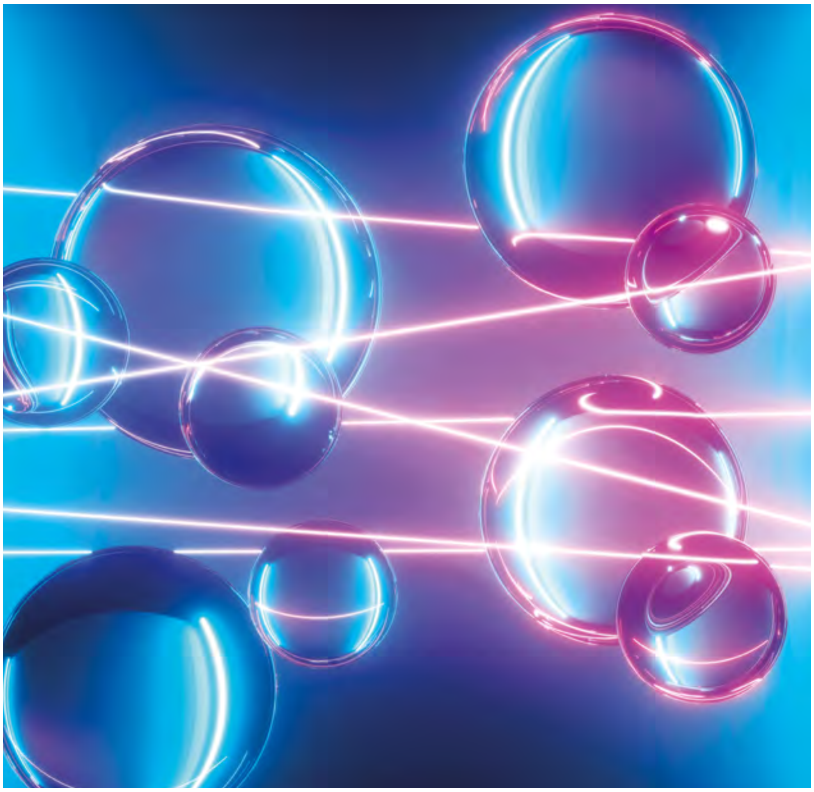
A pragmatist conception of competency in AI ethics

In order to think about AI ethics competency, we were quickly confronted with the fact that it was first and foremost important to

[1] A competency framework is a document that identifies and defines the competencies that a person taking a specific course should have developed by the end of it. This document is not a pedagogical tool, but a resource that can be used by administrations and program coordinators to revise or renew course offerings, or by teachers in designing a course outline. The framework can, however, serve as a model for the development of teaching materials and, as indicated later in the article, a pedagogical toolkit is currently under construction.

go back to the definition of ethical competency in general, AI ethics being a subset of ethical competency. However, this first exercise turned out to be more difficult than first thought, since we considered that certain pitfalls in training for ethical competency had to be avoided (Boudreau, 2019). On the one hand, ethical competency cannot be reduced to a discussion of the validity of classical frameworks in moral philosophy or to their simple application. Indeed, it seems that the most fruitful avenues in applied ethics avoid the application of classical ethical frameworks, which, as they are oriented toward the action of the moral subject, usually lead to a rather individual reflection, whereas in ethics in general, and even more so in AI ethics, moral problems are a matter of shared responsibility. Indeed, in our information societies, moral obligations cannot be understood only in their individual dimension, but, given the multiplicity of people involved in the design, development and use of technological devices, they must be understood collectively. We have thus sought to integrate the social and contextual dimension of ethics into our reflection. On the other hand, we also noted that another dominant approach in teaching applied ethics was characterized by a confusion between ethics and deontology, with ethics sometimes being reduced to an exercise of compliance with codes of ethics and codes of conduct in organizations (professional orders, businesses, public bodies, etc.) or with applicable laws. Such a conception of ethical competency is also unsatisfactory because it omits an essential component of moral posture: the reflective dimension. Indeed, ethical competency cannot be reduced solely to the ability to conform to pre-established rules but must include the ability to reflect critically and autonomously on the relevance of these rules.

Hence, it is with these reflections in mind that we were able to appreciate the specific contribution of a conception of ethical competency stemming from the philosophy of American pragmatism, in the wake of the work of C.S. Peirce, W. James and J. Dewey (Lacroix *et al.*, 2017; Keulartz *et al.*, 2002). The pragmatist approach to ethical competency is particularly interesting as it does not consider the individual in isolation, but rather presupposes a developing individual placed in a context of action and evolving as an integral part of social groups. This contextual approach to ethics addresses the two problems that have been discussed above. First, in such a perspective, it is not adequate to treat an ethical problem in an abstract and decontextualized way, which makes the pragmatist approach to ethics broader than the classical approaches to ethics teaching. On the other hand, the pragmatist approach to ethical competency poses from the outset the question of the validity and relevance of the principles that guide ethical action, thus avoiding the reduction of ethical action to an exercise of compliance. Such a reflexive approach to ethics does not involve applying a set of rules, but rather conducting an autonomous reflection on the principles and values that guide action (Boudreau, 2019). More specifically, the pragmatist approach to ethical competency has allowed us to identify three components of this competency (Lacroix *et al.*, 2017) which we have been able to translate into specific components of AI ethics in **Table 1**.



Source: Wacomka/iStock

Table 1

Definition of the three components of AI ethics competency

Being in an ethical situation *Ethical sensitivity*	To recognize and appreciate the ethical dimension of situations involving AIS in different spheres of our daily activities.
Knowing how to act in an ethical situation *Reflective capacity*	To problematize the ethical dimension of AIS issues and address these issues autonomously in order to act in an ethical situation.
Interacting in an ethical situation *Dialogic skills*	To state one's personal position on ethical issues related to AIS, evaluate the appropriateness of that position in comparison to other possible positions, and deliberate in order to coordinate joint actions with others in an ethical situation.

The four fields of competency in AI ethics

In the context of our research, we have found that the analysis of issues related to AI requires that we do not limit ourselves to ethical considerations in the narrowest sense of the term. The contextual analysis of ethical issues that we have found to be necessary for the definition of ethical competency requires in fact that we adopt a multidisciplinary approach that requires technical, philosophical, ethical, sociological and legal knowledge. In this sense, the mobilization of a pragmatist approach to ethics has led us to distinguish four fields of AI ethics competency in which the three components of ethical competency are deployed. These fields allow us to clearly identify the variety of sources of ethical issues specific to AIS. They include issues related to:

1. their technical functioning;
2. the specific moral dilemmas associated with them;
3. the socio-technical context in which they are embedded;
4. the complementary normative frameworks that guide their use.

1. Technical functioning of AIS

When it comes to establishing the ethical issues of AI, it is essential that this approach be based on a minimal understanding of the technical functioning of AIS. This allows us to focus on the real problems linked to concrete applications based on current uses of AIS. Furthermore, we believe it is also important to emphasize that such an approach should allow for an initiation to a reflection on the complexity of the relationship between humans and technology as well as on the impact of technology on our relationship to the world and on the construction of our personal identity, notions that belong to the field of the philosophy of technology.

2. Moral dilemmas specific to AIS

AIS replicate value conflicts that have been extensively discussed in the history of moral philosophy, such as the conflicts between transparency and privacy or between security and freedom of action. Nevertheless, AIS provide an opportunity for a new iteration of these conflicts, which requires that they be analyzed once again, possibly starting from redesigned conceptual frameworks. This may also give rise to new ethical issues, particularly in relation to the opacity of AIS and the associated requirements for explicability.

3. Socio-technical context of AIS

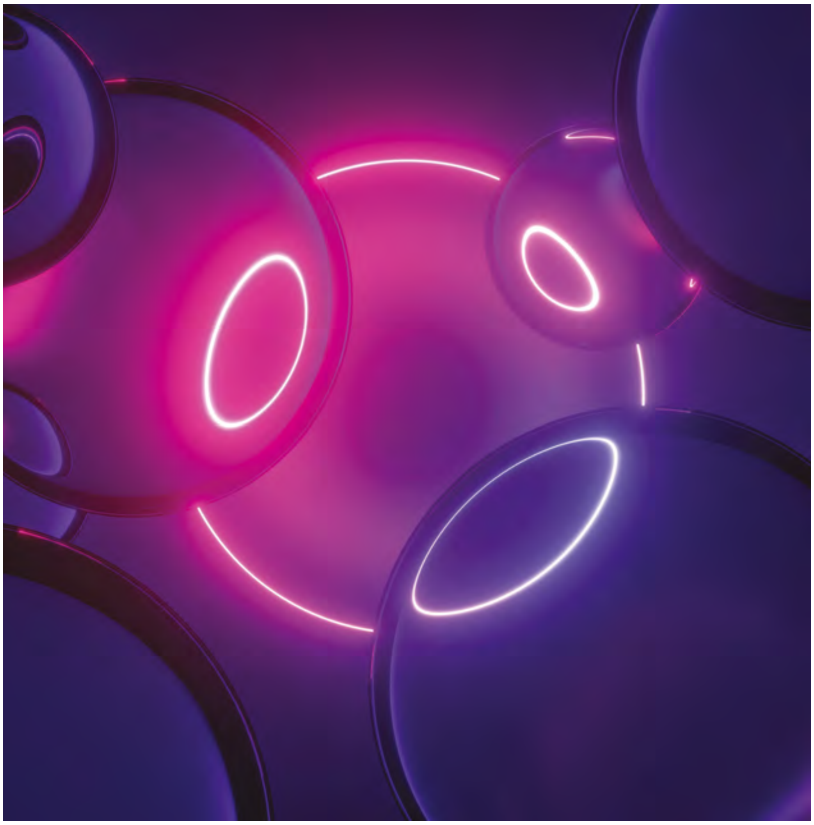
One of the important elements of the development of technological devices that allow the use of AI is linked to the socio-economic context of this development. Ethical issues are indeed part of a context marked by existing social and economic inequalities, inequalities that AIS cannot only reproduce, but also accentuate, by automating them. On the other hand, the ethical stakes of AI are also part of political dynamics that are important to grasp in order to put these issues in context. The same holds for the environmental issues related to the production and use of AIS.

4. Complementary normative frameworks related to AIS

Adequate and adapted training in AI ethics must include critical reflection on applicable laws related to AI, legislative reforms and their shortcomings, as well as other normative sources that frame AI, such as charters and ethical statements, but also other professional standards derived from ethical obligations and various codes of conduct.

In practice: the competency framework tested in a course

In the winter 2021 session, Cégep André-Laurendeau offered for the first time a new course entitled *Intelligence artificielle: l'être humain en transformation [Artificial Intelligence: The Human Being in Transformation*, Ed.], as part of the complementary course offer. This course, which aims to introduce students to the ethical and social issues related to AIS as well as to equip them adequately so that they can deal with these issues and consider possible solutions autonomously in order to resolve them, was designed based on the model proposed in the competency framework. The pedagogical approaches and teaching activities were also designed, in accordance with the pragmatist definition of ethical competency in the reference framework, to develop ethical competency in a contextual, practical, reflective and dialogical framework. While the application of the framework has been adapted to meet the requirements of a CEGEP complementary course, we believe that the document can lead to different pedagogical approaches depending on the teaching context. In the context of this complementary course, the subject matter has been divided into four thematic sections (see **Table 2**), based on the four fields of competency.



Source: Wacomka/iStock

Table 2

AI ethics competency applied to the complementary course *Intelligence artificielle: l'être humain en transformation*

Thematic section	In this thematic section, the students of the course were introduced to...	As an example, in this thematic section, students enrolled in the course were asked to...
Technical aspects of AIS	... the main characteristics of the technical functioning of AIS; ... different conceptions of the relationship between humans and technology.	... assess the degree of control they have over their own social networking practices (*ethical sensitivity*).
Moral dilemmas related to AIS	... the principles and values in conflict in the particular moral dilemmas specific to AIS; ... the various theoretical frameworks in moral philosophy that allow for the conceptualization of particular moral dilemmas related to AIS.	... identify moral dilemmas related to a fictitious situation involving the use of facial recognition in a shopping mall and deliberate on possible solutions to resolve them (*dialogic skills*).
Socio-technical context of AIS	... the social, economic and political context in which AIS is embedded; ... issues related to the environmental impact of AIS.	... measure the risks of discrimination associated with the use of AI for predictive policing (*reflective capacity*).
Complementary normative frameworks	... the various laws that govern the use of AIS; ... AI charters and ethics statements.	... critically assess the nature, scope and limitations of the new legislative provisions that frame the protection of personal data in Quebec and of the *Montreal Declaration for the Responsible Development of AI* (2017).

First, with the objective that they be able to target real issues related to concrete situations involving AIS and that they avoid associating the ethics of AI with scenarios that are more science fiction than reality, students were introduced to the history of AI, the different approaches and techniques for training AIS, and the ethical and social issues associated with these different practices, such as the black box problem, an expression that characterizes the difficulties generated by the high degree of opacity of AIS. In this segment, students were also asked to critically question the assumption of technological neutrality.

The second thematic section was designed to develop the ability to identify the value conflicts at stake in different types of situations involving AIS and, using the conceptual tools offered by different theoretical frameworks in moral philosophy, in the context of a dialogue between peers, to develop the ability to find ways to resolve them. Based on real or fictitious situations related to specific AIS, the participants had to mobilize the relevant conceptual resources in order to resolve the tensions between values such as privacy, security, equality or efficiency present in these situations, taking into account the point of view of all stakeholders as well as assessing the limits and inadequacies of the theoretical frameworks mobilized.

The third thematic section was dedicated to the specific issues related to the socio-technical context in which AIS are embedded. The effects of the asymmetric relationship between users and digital giants as well as the consequences of the prevalence of new business models based on the collection of personal data were studied. Issues related to the risk of reproduction, automation and accentuation of social inequalities by AIS, and more particularly the risk of bias and algorithmic discrimination, were addressed. In addition, issues arising from the environmental impacts of the increasing use of AIS in our various spheres of activity, including in our individual practices—from the extraction of ore required to build the devices that design and train AIS to the energy required to cool the servers that host the training data—were also studied.

The last thematic section of the course was designed to equip students with the knowledge of the different normative frameworks that govern the use of AIS and to allow them to position themselves critically in relation to these frameworks, whether they are legal, deontological or ethical in nature. In this segment, they had to mobilize these different frameworks and appreciate their nature, scope and limits in context.

At the end of this first exercise in the application of the competency framework—which, it should be emphasized, is not part of the PIA-funded research project but results from an independent process undertaken by the authors of this article—we were able to conclude that the students who took the course seemed particularly interested in such a pragmatist approach to AI ethics competency. The learners also responded positively to the pedagogical strategies geared toward the treatment of real or fictitious situations in which it is necessary to mobilize, at the end of an inquiry-like process, conceptual and dialogic resources in order to propose concrete solutions to problematic situations. We believe that such a contextual approach to AI ethics teaching—whose objective is to develop ethical sensitivity, the ability to problematize an ethical question in an autonomous manner in order to act on this situation to modify its course, as well as the ability to take into account the point of view of the various stakeholders in order to negotiate consensual solutions—allows for the adequate training of those enrolled. Thus, they will be able to deal with ethical issues related to AIS beyond the school context, in situations that will arise in their personal or professional lives or in the context of their civic activities. Given the rapid development of this technology and the necessarily novel nature of the problems to come, we believe it is important that education on AI ethics incorporate pedagogical strategies and learning activities that develop practical skills that can be deployed in different contexts.

Conclusion

In order to pursue these reflections, and in continuity with the model presented in the framework, the researchers have obtained new funding from the PIA that will allow them to design a pedagogical toolkit bringing together different turnkey activities that will be accessible online, and whose objective will be to develop the three components of the AI ethics competency in such a way that they can be deployed in the four fields of competency as defined in the framework. These activities will be designed so that they can be adapted to different teaching contexts. The pedagogical toolkit will be available online on Cégep André-Laurendeau's *bureau de la recherche et de l'innovation* (bRI) website at the end of the 2023 winter session. In sum, the pragmatist perspective that has been mobilized in the framework has given us the opportunity to reflect in greater depth on the value of this approach for ethics education at the college level. We think that if this reflection certainly extends beyond the scope of the projects discussed in this article, it would be relevant to develop it in another context in order to consider how this approach could be fruitful for thinking about ethical competency within the context of compulsory general education in philosophy. ▬

References

Boudreau, M-C. (2019). *La compétence éthique en milieu de travail : Une perspective pragmatiste pour sa conceptualisation et son opérationnalisation*, PhD thesis, Université de Sherbrooke.

Bruneault, F. , A. Sabourin-Laflamme and A. Mondoux (2022). *Former à l'éthique de l'IA en enseignement supérieur : Référentiel de compétence.*

Bruneault, F. and A. Sabourin Laflamme (2021). "Éthique de l'intelligence artificielle et ubiquité sociale des technologies de l'information et de la communication / comment penser les enjeux éthiques de l'IA dans nos sociétés de l'information," *TIC & société*, vol. 15, n° 1, p. 159-189.

Commission européenne (2021). *Proposition de règlement du parlement européen et du conseil établissant des règles harmonisées concernant l'intelligence artificielle et modifiant certains actes législatifs de l'Union.*

Conseil interprofessionnel du Québec (2021). *Présentation sommaire de l'encadrement actuel de l'intelligence artificielle.*

Gouvernement du Québec. Commission des partenaires du marché du travail. Ministère du Travail, de l'Emploi et de la Solidarité sociale (2022). *Se préparer à un marché du travail en transformation : Référentiel québécois des compétences du futur.*

Gouvernement du Québec. Ministère de l'Éducation et de l'Enseignement supérieur (2019). *Cadre de référence de la compétence numérique.*

Gouvernement du Québec. Ministère de l'Éducation et de l'Enseignement supérieur (2018). *Plan d'action numérique en éducation et en enseignement supérieur.*

Keulartz, J. *et al.* (2002). "Ethics in a Technological Culture. A Proposal for a Pragmatist Approach," in *Pragmatist Ethics for a Technological Culture*, Kluwer, pp. 3-21.

Lacroix A., A. Marchildon, and L. Bégin (2017). *Former à l'éthique en organisation.* Presses de l'Université du Québec.

PL 64, *Loi modernisant des dispositions législatives en matière de protection des renseignements personnels*, 42nd legislature, 1st session, 2021.

Tsamados, A. *et al.* (2021). "The Ethics of Algorithms: Key Problems and Solutions," *AI & Society*, vol. 37, p. 215-230.

Université de Montréal (2017). *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle.*

Andréane Sabourin Laflamme, M.A. (Philosophy), is a Philosophy teacher at Cégep André-Laurendeau. She is also a doctoral student in Law at the Université de Sherbrooke. Her research project, which focuses on the normative framework of AI, is funded by the Social Sciences and Humanities Research Council (SSHRC). She is a co-founder of the AI & Digital Ethics Lab (LEN.IA) and a college researcher at the International Observatory on the Societal Impacts of AI and Digital Technology (OBVIA). She contributes to various research projects in IA ethics and law.

andreane.sabourin-laflamme@claurendeau.qc.ca

Frédérick Bruneault, Ph.D. (Philosophy), is a Philosophy teacher at Cégep André-Laurendeau and associate professor at UQAM's École des médias. He is a co-founder of the AI & Digital Ethics Lab (LEN.IA), and a researcher associated with the Groupe de recherche sur la surveillance et l'information au quotidien (GRISQ), the International Observatory on the Societal Impacts of AI and Digital Technology (OBVIA) and the Observatoire du numérique en éducation (ONE). He leads several research projects in AI ethics.

frederick.bruneault@claurendeau.qc.ca