PART 1
CO-CONSTRUCTION APPROACH AND METHODOLOGY
CONTRIBUTORS

MARC-ANTOINE DILHAC, Scientific
Co-director of the Declaration, Full Professor,
Department of Philosophy, UdeM, Chair of
the Ethics and Politics Group, Centre
de recherche en éthique (CRE), Canada
Research Chair in Public Ethics and Political
Theory

CHRISTOPHE ABRASSART, Scientific
Co-director of the Declaration, professor
in the School of Design and Co-director
of Lab Ville Prospective of the Faculty
of Planning of the Université de Montréal,
member of Centre de recherche en éthique
(CRE)

TABLE OF CONTENTS

SUMMARY 25

1. INTRODUCTION 28

2. Why have a Montréal Declaration
for Responsible AI? 31
  2.1 The intellectual origins of this project 32
  2.2 Forum on the Socially Responsible
      Development of Artificial Intelligence 34
  2.3 Towards the Montréal Declaration
      for Responsible AI Development 35
  2.4 Montréal and the international context 36

3. THE ETHICAL AND SOCIAL ISSUES OF AI 38
  3.1 What is AI? 38
  3.2 AI in everyday life and philosophical
      questioning 39
  3.3 The Ethical issues of AI 41
  3.4 AI Ethics and the Montréal Declaration 42

4. THE CO-CONSTRUCTION APPROACH 44
  4.1 The principles of the co-construction
      approach 44
     4.1.1 The Principles of Good Citizen
          Involvement 44
     4.1.2 Experts and Citizens 46
  4.2 The Co-construction workshop methodology 47
  4.3 Uniqueness of the co-construction approach 48
  4.4 World cafés outside libraries 50
  4.5 Portrait of participants 50

This document is part of the 2018
MONTRÉAL DECLARATION FOR
A RESPONSIBLE DEVELOPMENT
OF ARTIFICIAL INTELLIGENCE.
You can find the complete report HERE.
TABLE OF CONTENTS

5. WORKSHOP DELIBERATIONS: EXAMPLES FROM SMART CITIES AND THE WORKPLACE 53
   5.1 The deliberation process 53
      5.1.1 Smart city sector: self-driving cars (SDC) and sharing the road equitably 54
      5.1.2 Workplace sector: Socially responsible restructuring? 58

6. PARTICIPANTS IN THE CO-CONSTRUCTION AND WORKING GROUPS 63

ANNEXES 68
   Annex 1 Co-construction workshops: Detailed description of how they work 68
   Annex 2 Foresight scenarios: winter co-construction workshops 70

CREDITS I
PARTNERS II

TABLE AND FIGURES

Figure 1: The Values of the Declaration (preliminary version) 31
Figure 2: The Values of the Montréal Declaration for a Responsible Development of Artificial Intelligence 32
Figure 3: The co-construction approach 35
Figure 4: Strategic forecasting: a three-step process 49
Figure 5: Proportion of men and women involved in the co-construction workshops 50
Figure 6: Participants in the co-construction workshops per age group 50
Figure 7: Distribution of participants in world cafés and co-construction days by education level reached 51
Figure 8: Distribution of participants in world cafés and co-construction days by field of activity 52

Table 1: Smart City, First deliberative moment: formulating ethical issues in 2025 55
Table 2: Smart City, Second deliberative moment: AI framework recommendations for 2018-2020 56
Table 3: Workplace, First deliberative moment: formulating ethical issues in 2025 60
Table 4: Workplace, Second deliberative moment: recommendations for an AI framework in 2018-2020 61
Table 5: Typical procedure for world cafés 68
Table 6: Typical procedure for co-construction days 69
Table 7: Scenario summaries 71
Table 8: Elements of five scenarios 73
On November 3, 2017, the Université de Montréal launched the co-construction process for the *Montréal Declaration* for a Responsible Development of Artificial Intelligence (*Montréal Declaration*). A year later, we present the results of these citizen deliberations. Dozens of events were organized to stimulate discussion on social issues that arise with artificial intelligence (AI), and 15 deliberation workshops were held over three months, involving over 500 citizens, experts and stakeholders from all backgrounds.

The *Montréal Declaration* is a collective endeavour that aims to steer the development of AI to support the common good, and guide social change by making recommendations with a strong democratic legitimacy.

The selected citizen co-construction method is based on a preliminary declaration of general ethical principles structured around seven (7) fundamental values: well-being, autonomy, justice, privacy, knowledge, democracy and responsibility. Following the process, the Declaration was enriched and now presents 10 principles based on the following values: well-being, autonomy, intimacy and privacy, solidarity, democracy, equity, inclusion, caution, responsibility and environmental sustainability.

If one of the goals of the co-construction process is to fine-tune the ethical principles suggested in the preliminary version of the *Montréal Declaration*, an equally important goal consists of making recommendations to provide a framework for AI research, as well as its technological and industrial development.

**First, what is AI?**

Very briefly, AI consists of simulating certain learning processes of human intelligence, learning from them and replicating them. For example, identifying complex patterns among a large quantity of data, or reasoning in a probabilistic fashion to sort information into categories, predict quantitative data or aggregate data. These cognitive skills form the basis for other skills such as choosing among several possible actions to reach a goal, interpreting an image or a sound, predicting a behaviour, anticipating an event, diagnosing a pathology and more. Two elements are key to these AI feats: data and algorithms, a series of instructions that perform a complex action.

To discuss the ethical issues of AI in concrete terms, the **co-construction workshop method** is based on the preliminary version of the *Montréal Declaration*. Schematically, after deciding on the “why?” (which desirable ethical principles should be included in a declaration on the ethics of AI?), it then becomes a matter of envisioning, along with participants, how ethical issues in the fields of health, justice, smart cities, education and culture, workplace and public services could arise in upcoming years. Then, we think about how we could respond to these issues. For example, through measures such as sector certification, a new stakeholder/mediator, a form or standard, public policy or a research program.

Citizens and stakeholders took part in world cafés or entire co-construction days where they had the chance to debate prospective scenarios.

Other citizens chose to contribute to the reflection by filling out an online questionnaire or submitting a paper. The results of these specific initiatives will be discussed in the global report on the activities associated with the *Montréal Declaration*, which should be published in the fall of 2018.

**SUMMARY**
Co-construction workshop results — General trends

In general, participants recognized that the advent of AI also brought important potential benefits. Participants especially recognized the time savings that AI devices could bring to their fields of work. However, they also felt that caution should be exercised in AI development to prevent it being abused or used for harmful purposes.

The citizens highlighted the need to implement different mechanisms to ensure that quality, understandable, transparent and relevant information was communicated. They also discussed the difficulty of guaranteeing truly enlightened consent.

The majority of participants recognized the need to align private and public interests, prevent monopolies from emerging and limit the influence of corporations.

Participants also recommended introducing mechanisms that would come from and involve people independently trained in technological and ethical issues of digital transition and AI to promote diversity, include the most vulnerable and protect the plurality of lifestyles.

No matter what it was used for, most participants insisted that AI remain a tool, and that final decisions be made by human beings when fundamental issues are at stake.

Priorities according to the principles of the Montréal Declaration

The principle of responsibility was considered the most pressing issue, followed by respect of autonomy and protection of privacy. Well-being, knowledge and justice came next. It should be noted, however, that they are all closely linked.

The principle of autonomy, considered a priority by most participants, entails respecting and promoting individual autonomy when they risk being controlled by technology and becoming dependent on tools. It also raises the issue of freedom of choice being two-sided: being able to make your own choice when faced with an AI-guided decision as well as the choice to not use these tools without risking social exclusion.

Participants also felt the principle of well-being was important. It was implicit at every roundtable, illustrating a collective desire to move towards a just and equitable society that fosters the development of all individuals. Overall, experts and users in every field concurred that the principle of well-being also serves as a reminder to maintain authentic human and emotional relationships.
Issues that led to creating new principles, or deliberating and exploring new themes

The principle of justice was discussed as the basis for two types of issues, which could lead to two new principles: a principle of diversity, which seeks to avoid discrimination by identifying bias-free mechanisms and a principle of equity or social justice, which states that AI benefits be accessible to all, and that its development not contribute to growing economic and social inequalities, but rather help bridge the gap.

A principle of caution. Issues related to trust in the development of AI technologies were regularly raised. This issue of trust is also closely tied to the question of reliability of AI systems.

A principle of explainability or justifiability. This principle implies being able to understand an algorithmic decision and react to it. For this, citizens felt it was important that algorithmic procedures be explained so they could see and understand which criteria were considered in the decision.

A principle of environmental sustainability. The impact of AI development and use on the environment raises specific issues, namely how to guarantee the responsible and fair use of material and natural resources.

Mechanisms for a digital transition

All the co-construction roundtables agreed on three (3) priority mechanisms to ensure socially responsible AI development, regardless of the field:

1. Include legal provisions.
2. Provide everyone with training.
3. Identify key independent stakeholders for AI management.

Pursuing the deliberations

The Montréal Declaration project focussed its first year of co-construction on many key sectors: education, health, work, smart city, predictive policing, environment, democracy and media propaganda. It is clear that a year of co-construction cannot possibly cover all the ethical and social issues associated with AI. The Montréal Declaration is not only the result of a collective reflection process, it is the very process itself: beyond Year 1 of the Montréal Declaration, the collective consultation and reflection process continues, because technological evolution waits for no one.

We present public policy recommendations around priority action areas. To date, four priority areas have emerged: algorithmic governance, digital literacy, diversity and inclusion, and environmental transition.
On November 3, 2017, the Université de Montréal, in collaboration with the Fonds de recherche du Québec, launched a co-construction process based on the Montréal Declaration for a Responsible Development of Artificial Intelligence (Montréal Declaration). We had no idea the level of interest this initiative would generate, nor of the size of the task that lay ahead. A year later, we present the results of citizen deliberations, which involved various groups from civil society, citizens, experts, professional bodies, industry stakeholders and policymakers. It was a resounding success: dozens of events were held to discuss social issues surrounding AI, and 15 deliberation workshops took place from February to October, involving over 500 citizens, experts and stakeholders of all professional backgrounds.

The report we are presenting must be taken as a summary of a democratic deliberation process to enlighten public policy decisions on artificial intelligence, an experience which can serve as a reference point for other deliberative forums. The work on what is called the Montréal Declaration was led by a multidisciplinary and inter-university team of researchers, mainly in Quebec but also across the world. Awareness of social issues around artificial intelligence is shared not only by this research community, but by society as a whole. We suggested a citizen co-construction process because we are convinced that everyone has a right to be heard about how our society should be. This approach is innovative in both content and form: first, because it introduces foresight methods of applied ethics, which consist of anticipating ethical controversies around future artificial intelligence technologies or social situations where these technologies are used in unprecedented ways. Following this, we carried out this consultation process on a vastly broader scope. The numbers mentioned above paint a clear picture. This process should continue beyond the public presentation of the Montréal Declaration, since it must remain open to review.

We solicited the public to draft the Declaration; in return, we were asked the following questions by not only the public, but various stakeholders: What will the Declaration change? Who is writing it? Isn’t this just a vain university endeavour? Aren’t there already too many manifestos, professions of faith on the ethical values of artificial intelligence? Isn’t developing artificial intelligence within a framework of ethical principles and recommendations a means of condoning it? Isn’t that approving a technocratic vision of society? Why not devote our energy to criticizing this development? None of these questions are without merit, and because we are committed to fostering greater transparency around artificial intelligence, we are also committed to increased transparency around the process we established. Our hope is that this report will provide a few answers.
The ethics of artificial intelligence have been a hot topic in many countries over the last two years. Each stakeholder in its development, not to mention researchers, businesses, citizens and political representatives, recognize the urgency of establishing an ethical, political and legal framework to guide the research and use of artificial intelligence. There is no doubt that with the rise of artificial intelligence technologies we are at the dawn of a new industrial revolution. The impact of this revolution on the production of goods, delivery of services, organization of work and the workforce, or even on family and personal relationships are still unknown but will be major, and possibly disruptive in certain fields. The social changes triggered by artificial intelligence are, indeed, surprisingly swift and illicit varied reactions, from enthusiasm to disapproval and scepticism. We could simply ignore them and embark in speculation about whether or not what we call artificial intelligence exists, but we'd only be postponing the problem to a point when it would no longer be possible to influence its development.

A number of objections and fears were raised during this first co-construction process. Many workshop participants and observers from the Declaration project questioned the technocratic ideology that views technology as a way of rationally organizing all of society, thus reducing social issues to technical problems. Others questioned the ability and the will of public institutions to regulate lucrative technologies. These objections cannot be casually dismissed, because they are based on historical precedents that shook our confidence in technological innovations, all the more so in the people promoting them. It is also important that individuals who raise objections do not undermine efforts to positively influence the future of society, but support them by getting involved in the democratic deliberations that allow us to maintain control over the development of digital technologies and artificial intelligence. We can complain about the effects of these new technologies on social relationships, or criticize how social life is being reduced to a series of lifestyles, but this will not prevent technological innovation, nor will it influence it. Yet that is the entire purpose of the Montréal Declaration: guide the development of artificial intelligence in order to promote fundamental ethical and social interests and provide guidelines for protecting human rights.

To conclude, we are not presenting a theory of artificial intelligence in this report, nor are we defending sophisticated arguments to settle the unrelenting question on the use of the term "artificial intelligence": is it an appropriate term to refer to data processing, recognition and decision-making algorithms? Some contest the use of this term by arguing that artificial intelligence refers to very limited knowledge processes when compared with human intelligence, or even the behavioural intelligence of pigeons. That is undeniable. But in that case, then, we must also recognize that complexity of paramecia surpasses that of any algorithm, even a learning one. If we go down that path, we will merely encounter roadblocks to understanding intelligence as a whole. What is human intelligence? Is there one or many forms? Do we need to introduce and specify an "emotional" form of intelligence? And in that case, why refuse to introduce an "artificial" form of intelligence? The hundreds of thousands of pages that have been written to answer these questions still do not suffice.

However, a few statements can help clear up misunderstandings that are at the root of the controversy: First, we know that the way biological neural networks operate is vastly different from that of artificial neural networks; there is no mistaking the two. But that does not invalidate the use of the term "artificial intelligence". If that were the case, the term mechanical arm would have to be discarded as well, given that a biological arm operates very differently, and that bones, joints, tendons and muscles are not pieces of metal, pulleys, springs and ropes. In general, people often confuse intelligence and thought. Intelligence is a property of thought, it is not thought as a whole. Intelligence, therefore, is particular in that it reduces the complexity of the world in which intelligent beings evolve to allow them to better master their environment. We give ourselves rules to analyze, calculate, evaluate and make decisions about reality. A long philosophical tradition of highly intelligent thinkers have asserted this, from Socrates to Russell to Leibniz. In a certain way, intelligence models and reduces reality to better act on it, like a mechanical equation models
and reduces movement to better understand it. Consequently, given the above, intelligence, even human, is largely algorithmic: it analyzes data and makes calculations according to procedures. It then lends itself to “mechanization” and “incarnation” very well, in the literal sense of the term: digital calculation, meaning calculations made on and with the fingers according to very diverse techniques, is an incarnation of calculations; with different abacuses such as the Chinese abacus, the Pascaline\(^1\) and the electronic calculator, we witness the mechanization of calculation.

Reflecting on the goals we wish to pursue is not strictly a matter of calculations. Building your personal and social life around certain worthwhile goals does not depend on an algorithmic function. Knowing if we must use nuclear weapons to kill the greatest number of people and weaken an enemy country cannot be solely determined by calculating the consequences; we must still define the good or the goods according to which the calculation of consequences has a moral sense. There’s something tragic about avoiding reflection on moral consequences by seeking only a calculation of the means. Artificial intelligence cannot yet engage in this kind of reflection. In the world we know and can anticipate in the near and mid-term future, reflecting on the finality of social life and existence in general is still a product of human intelligence.

The Montréal Declaration rests entirely on this statement: it is up to human and collective intelligence to define the purposes of social life and thereby, guide the development of artificial intelligence so that it is socially responsible and morally acceptable.

---

1 Mechanical calculator designed and presented by mathematician and philosopher Blaise Pascal in 1645.
The Montréal Declaration is a collective work that aims to accomplish three (3) goals:

1. develop an ethical framework for ethical AI development and deployment
2. guide the digital transition so everyone benefits from this technological revolution
3. create a forum for national and international dialogue to successfully pursue inclusive and equitable AI development.

It becomes, therefore, a question of using AI development to ensure everyone’s well-being, and guiding social change by developing recommendations founded in democratic legitimacy.

The Declaration is the outcome of an inclusive deliberation process that opens new dialogue between citizens, experts, public officials, industry stakeholders, civilian organizations and professional bodies.

The selected citizen co-construction method is based on a preliminary declaration of general ethical principles that is structured around fundamental values.

Our relationship to these “values” is then broken down into standards we call principles. For example, if the value is well-being, our relationship to this value is that of maximization: we must increase the well-being of sentient beings. If the value is autonomy, our relationship is that of respect or protection: we must respect the autonomy of moral beings. The goal of the initial task of identifying these values and principles was to launch a citizen participation process that would then define the ethical principles of responsible AI development and recommendations to implement to ensure that AI promotes fundamental human interests.

At the end of this process, the values and principles were fine-tuned, allowing us to pinpoint things more precisely:
Moving from a preliminary to final version of the Declaration resulted from discussions that arose during the public consultation and co-construction workshops. The choice of values and principles rests on an understanding of fundamental social expectations as they were expressed, and is motivated by a desire to examine priority issues and find balance between the different values for the sake of coherency. Because there is no template formula to select principles (no algorithm yet exists for this task), it is the result of a complex adjustment process generally referred to as deliberation.

2.1 THE INTELLECTUAL ORIGINS OF THIS PROJECT

The revolution in artificial intelligence (AI), and more specifically deep learning, opens our perspectives to unimagined technological developments that will help improve decision-making, reduce certain risks and help the most vulnerable. This revolution is remarkable in many ways, although it also brings up questions that were first raised in the 18th century during the Industrial Revolution. It would be unwise to ignore the unique aspect of this revolution by hiding behind platitudes that leave us ill-prepared to face current challenges. Of course, human beings are gifted beings with great technical abilities—human history is itself a history of technical transformations of nature, and artificial intelligence extends this trend to automation—but upon closer inspection nothing resembles what is at stake today with the arrival of artificial intelligence technologies. The cognitive skills we believed unique to humans can now be performed by algorithms, machines that must be recognized, in a certain sense, as intelligent.

Although the social impact of these new technologies is wide-ranging, it is still somewhat unknown. It could prove disastrous if we do not take the time now to think about the ethical, political, judicial, sociological or psychological ramifications on the type of society and human relationships we want to promote or protect while still benefiting from information technologies and algorithm calculations.

Using algorithms to make technical or administrative decisions is nothing new. Although algorithms have been around since the Middle Ages, the rise of decision-making algorithms truly began in the 1950s, especially in the field of healthcare: emergency room triage in hospitals, detection of sudden infant death syndrome risks, prediction of heart attacks. All these algorithm techniques—"procedures"—already raise a certain number of ethical and social issues: those

---

2 Algorithmic procedures have been known since Antiquity in fact, but contrary to what the “th” in algorithm may lead to believe, the word does not come from Ancient Greek, but rather from a Latinisation of the name of a mathematician living in Baghdad in the 9th century: Muhammad Ibn Musa Al-Khwarizmi. Latin translations of Al-Khwarizmi’s algebra manual had circulated throughout Western Europe as early as the 12th century, the first being the Cambridge manuscript Dixit Algorizmi. The original Arabian manuscript has been lost. Through distortion, al-Khuwārizmī thus became algorizmi and algoritmi, then algorithm. On the history of these texts, see André Allard’s reference edition, Muhammad Ibn Musa Al-Khwarizmi, Le calcul Indien (algorismus). Versions latines du Xlle siècle, Librairie scientifique et technique Albert Blanchard, Paris, 1992.
of social acceptability of an “automatic” decision, the final decision (is a human being at the end of the decision-making chain?), or responsibility in the event of a mistake. And, clearly, these issues are being raised again with the latest algorithmic innovations (see section 3 for a general presentation of artificial intelligence).

What is different, then, about the latest technologies that fall under the AI acronym? From an objective standpoint, changes include the quantity of information that can be handled by computers (big data) and the complexity of learning algorithms which, by feeding off big data, can perform perceptive and cognitive tasks that enable visual or audio recognition and make decisions in specific contexts. By combining different features (facial recognition, behaviour analysis, decision-making), AI raises extremely important ethical issues. From a subjective standpoint, what is new is the wake-up call to citizens, however late and suddenly this occurred, on issues of algorithmic governance, handling of personal data and the social impact already felt by some professional sectors.

If the progress of AI can surprise and fascinate, it can also evoke the fear that using machines, namely robots, will greatly diminish the aspect of human relationships when it comes to medical treatment, elderly care, legal representation or even teaching. Reactions to the development of artificial intelligence can even be hostile when AI is associated with increased control of individuals and society, a loss of independence and a curtailing of civil liberties. For this reason, a dark cloud always hangs over the hope that artificial intelligence will usher in social progress: placed in the wrong hands, AI could become a weapon of mass domination (control of private life, concentration of capital, new discrimination). Many people also question the intentions of researchers, developers, entrepreneurs and policymakers.

The development of AI and its applications therefore involve conflicting fundamental ethical values and create serious moral dilemmas and deep social and political controversies: should we promote public safety by increasing smart surveillance (facial recognition, anticipating violent behaviour) at the expense of individual freedoms? Can we objectively improve the well-being of individuals, namely by encouraging people to adopt behaviours normalized by smart devices (nutritional behaviour, work management, day planner) while still respecting people’s independence? Should economic performance targets take priority over a concern for an equitable share of the benefits of the AI market?

These dilemmas or tensions cannot simply be resolved by ranking fundamental values and interests. Otherwise stated: it is not about classifying values in order of importance a priori, or building a simple and unequivocal scale of values, let alone promoting some while ignoring others (security at the expense of liberty, efficiency without social justice, well-being at the expense of independence). We also cannot aspire to find unique and permanent solutions. What we need to do is seriously contemplate the moral dilemmas caused by the development of AI and build an ethical, political and legal framework together that will allow us to deal with this while respecting the different fundamental values that we legitimately hold as members of a democratic society.

2 Paul Meehl, Clinical versus Statistical Prediction, University of Minnesota, 1954.
2.2

FORUM ON THE SOCIALLY RESPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

These discussions were the starting point for an initiative by the Fonds de recherche du Québec and the Université de Montréal to organize an international meeting to discuss the social impacts of AI. Within this context, the Université de Montréal organizing committee suggested launching the work around the *Montréal Declaration for a Responsible Development of Artificial Intelligence* based on a consultative and participatory process⁴. On November 2 and 3, 2017, a forum on the ethical development of AI brought together leading experts in fields ranging from pure sciences to social sciences and humanities at the Palais des congrès de Montréal. The Forum proposed that guidelines for a collective reflection on the ethical and socially responsible development of artificial intelligence be established, with the following three objectives in mind:

> offer a public forum for dialogue on AI development issues and their social impact

> spark interest and raise visibility among decision-makers, industry partners, politicians and the general community interested in AI while bringing attention to social issues raised by the sudden growth and numerous uses of AI

> encourage an interdisciplinary and inter-industry approach as a key component to successful ethical and sustainable AI

That is how the guidelines were defined on an inclusive (interdisciplinary and inter-industry) approach, which is key to developing the *Montréal Declaration for a Responsible Development of Artificial Intelligence* that is responsible, socially progressive and equitable and promotes social justice. The preliminary version of the *Montréal Declaration* was presented at the end of the forum. It was then a matter of launching the citizen co-construction process on AI ethics, a process we will expand upon in section 33.

---

⁴ The Forum’s scientific committee was made up of Louise Béliveau (Université de Montréal, Vice-rectorat aux affaires étudiantes et aux études), Yoshua Bengio (Université de Montréal, Département d’informatique, MILA, IVADO), David Décary-Hétu (Université de Montréal, École de criminologie), Nathalie De Marcellis-Warin (École Polytechnique, Département de mathématiques et de génie industriel, CIRANO – Centre interuniversitaire de recherche en analyse des organisation), Marc-Antoine Dilhac (Université de Montréal, Département de philosophie, CRÉ Centre de recherche en éthique), Marie-Josée Hébert (Université de Montréal, Département de philosophie et CRÉ – Centre de recherche en éthique), Catherine Régis (Université de Montréal, Faculté de droit, CRDP – Centre de recherche en droit public), Christine Tappolet (Université de Montréal, Département de philosophie et CRÉ – Centre de recherche en éthique).
2.3
TOWARDS THE MONTRÉAL DECLARATION FOR RESPONSIBLE AI DEVELOPMENT

Figure 3: The co-construction approach

As mentioned at the beginning of the chapter, the initial work of identifying these values and corresponding principles was conducted solely to launch the citizen participation process so that the ethical principles of responsible AI development could be refined, added to and completed. It should come as no surprise, then, that the preliminary version of the Declaration is schematic and that the statement of principles is intentionally very simple and consensus-based, so that they could be interpreted and completed during the public deliberations. One year later, the Declaration has been considerably improved.

The scientific committee in charge of writing this preliminary version was made up of Yoshua Bengio (Université de Montréal, Département d’informatique, MILA, IVADO), Guillaume Chicoisne (IVADO), Marc-Antoine Dilhac (Université de Montréal, Département de philosophie, CRÉ – Centre de recherche en éthique), Vincent Gautrais (Université de Montréal, Faculté de droit, CRDP – Centre de recherche en droit public), Martin Gibert (CRÉ – Centre de recherche en éthique, IVADO), Pascale Lehoux (Université de Montréal, ESPUM – Ecole de santé publique), Joëlle Pineau (Université McGill, School of Computer Science; MILA), Peter Railton (Université du Michigan, Académie américaine des arts et des sciences, philosophie), Christine Tappolet (Université de Montréal, Département de philosophie et CRÉ – Centre de recherche en éthique).
If one of the goals of the co-construction process was to refine the ethical principles proposed in the preliminary version of the Montréal Declaration, another equally important goal was to develop recommendations for overseeing AI research and its industrial and technological development. However, all too frequently we see analysis reports and recommendations forgotten as soon as they are published; that is why we must keep the momentum going during the co-construction period.

Once the co-construction process is complete (or suspended), we need to open public debate in forums where political, legal and policy decisions are made, so that recommendations from citizen deliberations may be concretely implemented. These recommendations are not simply legal in nature and, when they are, do not necessarily involve changing a law. They could, however, demand that the legal framework be modified; in some areas, they must. In other instances, the purpose of the recommendations is to feed and guide discussions held by professional organizations so that they modify their code of ethics or so that companies adopt a new ethical framework.

This step is the ultimate goal of the co-construction process. We must, however, immediately clarify that, when faced with a technology that has not stopped evolving over the past 70 years and whose major innovations arise every two to five years on average, it would be unreasonable to present the Declaration as definitive and complete. We need to think of co-construction as an open process, with successive and cyclical stages of deliberation, participation and recommendations, and see the Declaration itself as a road map that can be reviewed and adapted as AI knowledge and techniques evolve. This process of knowledge production, citizen deliberations and ethical framework and public policy recommendations will need to be expanded to a lasting institutional structure that allows it to respond to the evolution of AI.

2.4

MONTRÉAL AND THE INTERNATIONAL CONTEXT

The Montréal Declaration initiative is part of a dynamic scientific, social and industrial context. Montréal is a major hub for research and development in artificial intelligence, boasting a community of researchers, world-renowned university labs (MILA, IVADO) and an incubator full of thriving start-ups and businesses. This scientific, technological and industrial development is at the heart of a revolution that is transforming social practices, business models and lifestyles, and affecting all sectors of society. Thanks to its Laboratoire de l’innovation urbaine de Montréal, the Ville de Montréal is also a living lab of social and technological change. With fundamental scientific research come social and ethical responsibilities that the Montreal AI community fully embraces.

But outside Montréal, Quebec and Canada also offer a social context that is conducive to reflecting on the social impact of AI. Like MILA in Montréal, Vector in Toronto, AMII (Alberta Machine Intelligence Institute) in Edmonton, and the CRDM (Centre de recherche en données massives) in Quebec are hubs of excellence in fundamental research that have brought about incredibly quick and robust industrial growth. The Canadian Institute for Advanced Research (CIFAR, or ICRA, a partner in the Declaration project) has played a leading role in the Canadian development of AI by supporting fundamental research when AI was going through its “winter”. The Declaration initiative is supported by various stakeholders in Québec and Canada outside of Montréal.

Many international partners have also shown their support for the Montréal Declaration, especially its methodology. The Declaration team was able to establish a dialogue with institutions such as the Royal Society of the United Kingdom, the EGE (European Group on Ethics in Science and New Technologies) and the European Commission’s HLEG (High Level Expert Group on AI), which have

6 http://ville.montreal.qc.ca/portal/page?_pageid=5798,141982209&_dad=portal&_schema=PORTAL
7 We wish to thank UK Science and Innovation Network in Canada who facilitated the dialogue.
8 The European Group on Ethics in Science and New Technologies (EGE) is an independent advisory body of the President of the European Commissions.
their own study program and recommendations on AI. Immediately, we note similarities in the guidelines for ethical AI development as well as a shared desire to promote a democratic notion of AI use for the common good.

The Montréal Declaration initiative must also be viewed in the international context of an AI spring. The numerous initiatives that came before it must be highlighted because they acted as a catalyst for discussion on responsible AI. First, the Future of Life Institute, which was created in 2014, drafted the Asilomar Declaration in 2017: following a three-day conference, a declaration containing 23 fundamental principles on AI research and its uses was signed by more than 1,200 researchers. Professor Yoshua Bengio took part in the event and brought attention to the risks of irresponsible and malicious AI use.10

Since the Asilomar Conference, there have been many reports published on AI ethics. The report from the Institute of Electrical and Electronics Engineers (IEEE), Ethically aligned design V2, was made public at the end of 2017 and brought together several hundred AI researchers and engineers. The AI Now Institute based in New York University has also produced several reports, the latest of which deals with evaluating the impact of AI.11 Two ambitious strategic reports were published in March and April of 2018: the Mission Villani report in France and “AI in the UK: ready, willing, and able?” from the United Kingdom House of Lords. We must also highlight the participative approach of the CNIL (Commission nationale de l’informatique et des libertés) in France that led to the publication of a report with the evocative title, “How can humans keep the upper hand? - The ethical matters raised by algorithms and artificial intelligence”, in December 2017.

How does the Montréal Declaration position itself among these many independent initiatives? And what about the rise of ethical issues in AI? This last question so important that we include the same warning as the EGE in its report Artificial Intelligence, Robotics and “Autonomous” Systems (March 2018) that in the absence of coordinated reflection on the ethical and social issues of AI, a risk of “ethics shopping” exists. The immediate consequence is a sort of “offshoring” of ethical costs in areas of the world where ethical criteria are low priorities. Another risk is trivializing ethical discourse.

Each step in developing the ethical framework has merit. Part 2 of this report provides a “Overview of international AI ethics recommendations for 2018”. The Montréal Declaration initiative is different in that it is essentially participative. From February to November 2018, the co-construction process brought together over 500 citizens, experts and stakeholders during 15 workshops, co-construction days and roundtables across Québec and Europe. Although other participative initiatives have been led elsewhere, namely in France, the Montréal Declaration is unique in both its size and foresight methods.

The Montréal Declaration’s vocation is to open a forum for dialogue in Québec and Canada and offers a platform for a collective think tank that extends beyond Canadian borders. The goal is to identify socially acceptable and innovative AI trends using informed citizen discussions as a benchmark for the different democratic societies concerned. Citizens of non-democratic societies who wish to take part in a global debate on the future of human societies must also have access to this forum for dialogue.

---

9 The HLEG on AI is a group of 52 experts selected by the European Commission to define the application principles of Europe’s AI strategy. We thank the people in charge of the HLEG for allowing us to take part in their work between September and November 2018, in order to share and enrich our respective reflections and experiences.

10 Yoshua Bengio interview during the Asilomar conference: futureoflife.org/2017/01/18/yoshua-bengio-interview/


3. THE ETHICAL AND SOCIAL ISSUES OF AI

The collective reflection process at the heart of the development of the Montréal Declaration is based on a preliminary version of the Declaration of ethical principles itself and informative exposés on AI and the ethics of AI.

3.1 WHAT IS AI?

The idea of AI is not new. As early as the 17th century, philosopher and mathematician Leibniz came up with the idea of a universal characteristic and combinatorial art: reasoning comes down to calculating, and thought is conceived in algorithmic fashion. The notion of calculus ratiocination (logical calculation) predates the idea of an intelligent machine as it was developed three centuries later in the 1940s by Alan Turing. In a 1948 report entitled "Intelligent Machinery" and in 1950, in his famous article "Computing Machinery and Intelligence" Turing discusses a machine’s intelligence and develops the imitation game to define the conditions in which a machine can be said to think. The term artificial intelligence appears for the first time in 1955 in the description of a workshop offered by John McCarthy (Dartmouth College), "2-month, 10-man study of artificial intelligence". But the uses and development possibilities seemed very limited then, and so began the AI winter, with minimal interest from the scientific community. Yet, if the discipline’s development paled in comparison to the philosophical and cultural fervour it inspired (one need only recall 2001: A Space Odyssey, Blade Runner or Terminator, to name but a few hit movies), research in the field never ceased, and the dawn of the 21st century ushered in an AI spring.

In a certain way, AI consists of simulating human intelligence, drawing inspiration from it and reproducing it. But, above all, it is the brain, the human intelligence headquarters, which was designed as a machine capable of gathering, identifying and collecting data from its environment that it can then analyze, interpret and understand, using this experience to establish connections. The field of AI research consists of producing mathematical tools to formalize how the mind operates, thereby creating machines that can perform more or less basic cognitive tasks associated with natural human intelligence. For example, recognizing complex patterns among a large quantity of data, or reasoning in probabilistic fashion to classify information according to categories, predict quantitative data or group data together. These cognitive skills are the basis for other skills such as deciding among many possible actions to achieve a goal, interpreting an image or sound, predicting behaviour, anticipating an event, diagnosing a condition and so forth.

But these cognitive skills can only exist if the machine is also capable of identifying sensitive shapes such as images and sounds, which has been made possible by recent computer innovations. The notion of AI, therefore, also encompasses visual or sound recognition technologies that allow the machine to perceive its environment and construct a rendering of this environment.

Two elements underpin the achievements of AI: data and algorithms, meaning a series of instructions that perform a complex action. Simply put, if you want to cook a new dish, you need to know the ingredients (the data) and follow a recipe that provides instructions on how to use them correctly (the algorithm). Up until now, data processing capacities

---

13 Leibniz (1666), De Arte combinatoria.
15 Alan Turing begins his “Intelligent Machinery” (1948) report as follows: “I propose to investigate the question as to whether it is possible for machinery to show intelligent behaviour.”
(quantity of data and processing algorithms) were too limited to imagine a useful development for AI technologies. Things changed with the use of materials that made it possible to build very small and very fast calculators (computer chips) and store massive amounts of data as well as the dawn of the information era with the Internet.

What changed is the gigantic amount of data we can not only generate and transmit, but also process. If big data existed in the past, for example in the financial industry, nowadays it is a multitude of inanimate objects, spaces or receivers that are constantly producing unstructured or structured data, which must be manipulated and transformed before it can be used (data mining). It can be millions of messages published on social media, all the words contained in a library full of thousands of books, or content from huge image banks.

But what changed also is the type of algorithm developed by AI researchers. Determinist algorithms, which are a determined set of instructions like a cooking recipe, are being replaced by learning algorithms which rely on increasingly complex neural networks as the calculating power of machines increases. In computing, we talk about machine learning, and the progress of this field of study was reinforced by the development of deep learning. At the heart of the notion of AI itself is the ability to adapt and learn. In fact, for a machine to be considered intelligent, it must be able to learn by itself from the data it receives, as a human being does. And just like with humans, machine learning can be supervised, or not, by human beings that train machines on data.

It is these deep learning techniques that allowed machines to surpass human beings in complex games such as chess with AlphaZero, which also beats any other machine that doesn’t use deep learning, and the game of Go, which was reputedly unbeatable at algorithms, but which saw AlphaGo triumph over the best players in the world in 2015.

Although these examples are telling, AI can also serve other purposes such as automating tasks that require human intervention, especially tasks such as perception and recognition. For example, processing speech, recognizing objects, words, shapes and text, interpreting scenes, colours, similarities or differences in large sets, and by extension analyzing data and decision-making—or help with decision-making. The possibilities are incredibly vast, and increase tenfold every time engineers and programmers combine them to create new uses.

3.2

AI IN EVERYDAY LIFE AND PHILOSOPHICAL QUESTIONING

AI engages us in an ethical reflection that, unlike one concerning nuclear or genomics, deals with everyday objects and technologies. AI is all around us and shapes our lives more than ever. We are used to wearing small connected objects (phones, watches) and we are preparing for the arrival of self-driving vehicles, cars and buses, but already we take trains and subways that operate independently, and planes on autopilot can take off, steer and land without human intervention. We use ranking algorithms for our Internet searches, autocorrect built into our messaging apps, curation apps for music or meetups, and we know that companies use sorting algorithms, banks use management and financial investment algorithms, and that certain medical diagnoses can now be very exactly made by algorithms, and the list goes on.

These technologies are so seamlessly integrated into our everyday life that we no longer really think about them. When we talk about AI, most people still associate it with menacing, multifunctional machines that have some sort of consciousness, able to formulate a plan to destroy all humans. Yet the AI experience is a thoroughly banal one nowadays, with recommendation algorithms flooding the Internet (Google, Amazon, Facebook). If you’re shopping online, there’s a good chance a pop-up window will open and that Inès will start up a conversation with:

“Hi, how can I help you shop today?”

“Hi Inès”

Stanley Kubrick masterfully captured (and helped craft) this fantasy with the very human computer HAL 9000, in his film 2001: A Space Odyssey (1968).
For a few moments, you have the impression that there is someone named Inès behind the screen talking to you; for a few moments, it’s okay to wonder. Inès asks you questions, answers yours, provides the important information you need to continue shopping. But after a little back-and-forth, you realize that although Inès provides relevant information, she replies in mechanical fashion, she doesn’t understand the way you write, doesn’t get jokes or open-ended questions, in other words, she doesn’t interact naturally with you. Inès is a conversational agent, a chatbot, AI. It has become commonplace to chat online with chatbots to get more information about your health plan or new bank account, or even fashion advice.

For now, chatbots can be spotted within a few minutes of conversation, usually much sooner. If a chatbot could go undetected by a human being for a reasonable amount of time, we would consider that the machine successfully passed the Turing test and we would, according to this test, be dealing with a case of artificial intelligence, meaning a machine that thinks.

In his famous article, “Computing Machinery and Intelligence”, the father of modern computing, Alan Turing, proposes an answer to the question: “Can a machine think?” And yet, in the introduction of his article, he changes the problem he feels he can provide an answer to: can a machine act in such a way that it is indistinguishable from a human being? He then offers the famous “imitation game”, which consists of putting a human being that asks questions (the interrogator) in contact with another human being and a machine answering his questions. If the machine can imitate a human being to the point where the interrogator cannot tell whether the human being or the machine replied, we can conclude that the machine thinks. This is what is meant by the “Turing test”.

This imitation game caused a great deal of controversy and saw philosophers fiercely opposed over whether a machine could be said to think. An experiment known as the “the Chinese chamber” was made popular in the 1980s by philosopher John Searle. According to Searle, a machine that outwardly acts in the same fashion as a human being cannot be considered to have intelligence in the true sense of the word. To illustrate this point, Searle asks us to imagine a room in which a person who, knowing nothing of Chinese, will try to pass for a Chinese speaker. It’s a variation of the imitation game: the person in the Chinese room, let’s call him John, receives messages written in Chinese that Chinese speakers outside the room hand him. John doesn’t understand a word of the messages he receives, but he possesses a very complex instruction manual which allows him to manipulate the Chinese characters and compose replies that are understood by Chinese speakers outside the room, so that they believe that the reply was written by someone who speaks Chinese. Searle deducts that in this case John simulated language skills but doesn’t possess them; he made people believe he understood Chinese, but he didn’t understand what he was writing. According to Searle, the same conclusion goes for AI: an intelligent machine manipulates characters, it follows an algorithm, meaning a series of instructions to accomplish a task (in this case, write), but doesn’t understand what it’s doing.

The debate is a fascinating one and far from being settled, but we don’t really need to answer Turing’s question to consider the place AI holds in our lives and in our societies. For now, well-trained chatbots can converse as well as humans within a very limited framework of conversations, but leave no one guessing once that framework changes. And even if AI is ushering in an era where it is harder and harder to tell a naturally intelligent being from an artificially intelligent one, intelligent machines remain tools developed to accomplish well-defined tasks. We can, then, leave it up to cognitive philosophy metaphysics, psychology and neuroscience to debate the concept of artificial intelligence and discuss the possibility of robots developing emotions and feeling empathy.

The questions that arise with the introduction of AI in our lives are of a practical nature, whether ethical, political or legal. It is a questioning of the values and ethical principles, public policy orientations and applying standards surrounding AI research and its uses.

17 A. M. Turing (1950).
19 Which is very different from questions on the use of machines to detect human emotions, process them and answer in adequate fashion. See for example the work of Rosalind W. Picard, Affective Computing, Cambridge, MIT Press, 1997.
Because AI technologies are indifferent to their multiple uses, the problem is not knowing whether AI is good or bad in and of itself, but of determining which uses and goals are ethical, socially responsible, and compatible with democratic values and political principles. However, this ethical reflection does not only concern the uses of AI, but also AI research, its general orientations and goals. Nuclear research was not initially destined to produce bombs with tragically powerful consequences for humanity. Many scientific programs, however, did have that goal. We must, then, pay close attention to the direction AI research takes, in universities and as well as private corporations or government organizations.

3.3 THE ETHICAL ISSUES OF AI

Why introduce ethics when we can discuss the societal, social and economic impacts of AI? Can we afford the luxury of an ethical reflection? And isn’t it a bit naive to want to provide an ethical framework for AI development, which generates colossal profits? These are questions ethicists hear on a regular basis among sceptical citizens, as well as decision-makers who understand the extent of their leeway. To answer this, we must first briefly present the field of ethics when discussing the social issues of AI.

Simply put, ethics is a reflection on the values and principles that underlie our actions and decisions, when they affect the legitimate interests of other people. This supposes that everyone can agree on a person’s legitimate interests, and this is precisely what feeds the debate in ethics. The field of ethics is therefore not concerned with what can be done, but generally what must, or should be done: we can kill a million people with a single nuclear bomb, but must it be done to impress an enemy country and demoralize a population already suffering from war? Take a less tragic example: you can lie to a friend about their new haircut, but is it moral in order to save them from being hurt? What must be done in this case? To answer that question, we must examine the available options: tell the truth, or not tell it, or tell only part of it, or tell it in a certain way. We must also examine the consequences of each option, question if they are important, and if so, why. We must also reflect on the objectives which are valorous (doing good unto others, respecting others). Finally, we must give ourselves a rule, a moral principle: for example, the categorical principle according to which it is always wrong to lie, regardless of the consequences; or the hypothetical principle according to which it is not morally right to lie unless...

The field of ethics that applies to AI issues is public ethics. If we use the same type of reflection as public ethics, the subject isn’t the same, nor is the context for reflection. Public ethics is concerned with all questions that involve difficult collective choices on controversial institutional and social practices that affect all individuals as members of society, and not as members of a particular group: should a doctor tell his patient the truth about his health condition even if it will depress him and speed up the progress of the disease? This question doesn’t concern the doctor’s personal morality, but the types of behaviour we can rightfully expect from someone who holds the social role of doctor. This question is of a public nature and should be the subject of a public debate to define, using social values, best practices in terms of the patient–doctor relationship. By public debate, we mean all types of discussions which can take many forms of consultations, deliberations or democratic participation, and which are open to diverse individual and institutional stakeholders such as professionals in the field, association or union representatives, experts, policymakers and citizens. Public ethics calls for a collective reflection to establish best practice principles and demands that the stakeholders justify their suggestions based on acceptable arguments in a context of pluralism. In the case of the medical lie, you can appeal to shared values such as independence, respect for the person, dignity, the patient’s health or well-being, etc. Out of these values, principles can be established that guide the practice of medicine and provide paths to regulation by implementing a code

Public ethics are not beside nor above the law, which has its own logic, but help clarify the issues of social life that various stakeholders must keep in mind to meet citizens’ standard expectations and ensure equitable social cooperation. In this sense, public ethics shape public policies, and can lead to legislation, regulation, a code of ethics, an audit mechanism and more.

In AI, it is this kind of ethical reflection that we introduce. Let’s take the example of Melody, a medical conversation agent. Melody makes online diagnoses that you can access on your cellphone according to the symptoms you describe. In a certain way, it acts like a doctor. This can be very practical in a society where the healthcare system is either inaccessible or underdeveloped. But the fact that it is practical is not sufficient to authorize the public release of an app like Melody. Indeed, this app raises ethical questions that were not readily apparent with Inès, the shopping advisor chatbot. For example, we need to debate whether Melody must give users every possible prognosis, even if they are not equipped to understand the information. This problem simply transposes ethical medical questioning which has already received a normative response for which there is widespread consensus. The notion of informed consent, of a patient’s free and enlightened decision helps clarify a doctor’s obligations. Does this solve the problem that Melody and its sister applications often multiply unchecked? Overall, probably, but when specific attention is paid to this technology it is not that simple. The context does not allow Melody to ensure that the patient understands the diagnosis, or the urgency or not of treating the diagnosed condition. What rules must be invented to guarantee a patient’s autonomy and well-being? That is the issue of collective deliberation on AI’s ethical issues.

Ethical solutions have yet to be found for other issues specific to AI. For example, if Melody makes a wrong diagnosis, and the condition of the user who followed her advice takes a serious turn for the worse, who is responsible? In the case of a medical consultation with a human doctor, it is very easy to determine who is responsible for a medical error, but that is not the case with decision-making algorithms. Do you hold the algorithm responsible? The developer, or rather the company that developed the algorithm and that makes money from its use? And if the product is certified, shouldn’t the certifying body be blamed and penalized?

Public ethics questioning forces us to think about institutions that offer credible responses to a moral dilemma. It also deals with the type of society we want and the principles of its organization. By pursuing the reflection on medical chatbots, the use for developing such intelligent machines, from a social and human standpoint, is undeniable. We must question, then, whether it is acceptable for smart apps to replace medical doctors, assuming they can make an accurate diagnosis, even more accurate than a human. What does a patient-doctor relationship look like when the doctor is a chatbot? What essential elements are gained and which are lost? This is not a “utilitarian” type of question, but a question about the importance of our social relationships, recognizing our vulnerability as patients, our human identity. Let’s take it one step further: investing in the development of this type of AI rests on an eminently arguable social choice, which requires a collective discussion on the type of society we wish to build. We can consider the need for improving access to an efficient public healthcare system and therefore further invest in medical training and an equitable health organization.

3.4
AI ETHICS AND THE MONTRÉAL DECLARATION

The development of AI and its uses, then, involves fundamental and conflicting moral values that can provoke serious ethical, social and political controversies: should we develop apps like Melody to diagnose isolated people more quickly, or invest differently in the healthcare system so everyone can see a human doctor? There is no simple answer, but choices must be made.

---

20 The British public health service, the NHS (National Health Service) recently created a library of trustworthy apps (NHS Apps Library). Apps that do not offer sufficient guarantees can be deleted from the library, which brings serious commercial repercussions for the company selling the app.
The *Montréal Declaration* provides a basic moral vocabulary to identify, analyze and form practical answers to problematic social situations. The analysis of the Melody chatbot case illustrates the purpose of the Declaration. To understand the issue of enlightened patient understanding of a diagnosis, attributing fault in the event of a wrong diagnosis or accessing health services, the *Montréal Declaration* offers a list of values you can immediately consult: autonomy, responsibility, equity or justice. The principle of privacy, for example, helps frame the problem of patient data confidentiality.

The Declaration’s first objective consists of identifying the ethical principles and values that promote the fundamental interests of people and groups. When applied to the field of digital and artificial intelligence, these principles remain general and abstract. To understand them properly, it’s important to keep the following elements in mind:

1. Although they are presented as a list, there is no hierarchy. The last principle is no less important than the first. However, depending on the circumstances, it is possible to lend more weight to one principle than another, or to consider one principle more relevant than another.

2. Although they are diverse, they must be subject to a coherent interpretation to avoid any conflict that could prevent them from being applied. As a general rule, the limits of one principle’s application are defined by another principle’s field of application.

3. Although they reflect the moral and political culture of the society in which they were developed, they comprise the basis for an intercultural and international dialogue.

4. Although they can be interpreted in different ways, they cannot be interpreted in just any way. It is imperative that their interpretation be coherent.

5. Although these are ethical principles, they can be translated into political language and interpreted in legal fashion.

The Declaration of principles is followed by a list of recommendations that act as guidelines for the digital transition within the Declaration’s ethical framework. This list does not aspire to be exhaustive, nor can it cover every aspect of AI application; such ambition would be doomed. Rather, it aims to address a few key cross-sector themes so that we can think about how to make the transition towards a society in which AI helps promote the common good: algorithmic governance, digital literacy, digital inclusion of diversity and environmental sustainability.

The *Montréal Declaration* was designed for any person, organization and company that wishes to take part in the responsible development of artificial intelligence, whether to contribute scientifically or technologically, develop social projects, establish rules (regulations, codes) that apply to it, contest harmful or unwise approaches, or alert public opinion when necessary.

It is also designed for political representatives, whether elected or appointed, whose citizens expect them to respond to developing social changes, quickly establish a framework that encourages digital transition for the greater good, and anticipate serious risks presented by AI development.

The recommendations that follow the Declaration are intended more specifically for stakeholders in AI development in Quebec and Canada. They are examples of concrete measures collectively developed out of the Declaration’s ethical considerations. For this reason, they can act as a benchmark for stakeholders in AI development outside Canada.
4. THE CO-CONSTRUCTION APPROACH

4.1 THE PRINCIPLES OF THE CO-CONSTRUCTION APPROACH

To answer the many questions raised by the use of intelligent machines and ensure that AI develops "in an intelligent manner" within democratic societies, we need to solicit an "excess" of democracy and involve the greatest number of citizens in the reflection process on the social issues of AI. The goal of the co-construction approach is to open a democratic discussion on the way society must be organized to use AI responsibly.

It is not only a matter of knowing what people think of a certain innovation and surveying their "intuitive" preferences; co-construction is not a public opinion poll on questions such as: "Are you scared that AI will replace judges?", "Would you prefer that a human or a robot operate on you?" This is an interesting question, and the survey method provides important information to policymakers as well as important working material for social sciences. However, although co-construction invites people to think collectively about democratic issues, it also calls for well-argued answers to pressing questions to be developed and political and legal recommendations to be formulated. The co-construction process also lends them a certain democratic legitimacy, which creates the conditions for a political debate and accountability from policymakers, professionals and industry stakeholders.

This is the entire reasoning behind the approach initiated by the Montréal Declaration: entrust democratic societies with the responsibility of resolving moral and political issues that affect society as a whole. The future of AI is not only written in algorithms; it resides foremost in collective human intelligence.

4.1.1 The Principles of Good Citizen Involvement

The moment you involve the public in a consultation and participation process (co-construction) on controversial social questions, you must ensure that the process avoids the risks usually associated with a democratic exercise. And yet, two objections are traditionally brought up to discredit public involvement:

1. Ignorance: according to this objection, the most common, the public is ignorant and cannot understand complex issues that require scientific knowledge, master logical forms of argument and understand political and legal processes.

2. Populism: according to this objection, involving an unqualified public creates an opportunity for demagogic manipulation that fuels popular stereotypes and can lead to unreasonable proposals being adopted that are hostile to social progress or tyrannical towards minorities.

We do not share the belief that the public is so ignorant that they must not be consulted. We do not subscribe to the idea that non-expert members of our society have unsurmountable prejudices and their alleged irrationality causes them to make systematic errors. Ignorance is certainly an important problem, but we believe instead that they can shed light on neglected aspects of social controversies, because they are concerned by the issues discussed, and they can contribute to significant solutions that experts haven not thought of, or were unable to support publicly.

If, for some individuals, prejudices and irrational tendencies cannot be completely eliminated, these biases can be overcome collectively. In favourable conditions, non-expert individuals can take part in complex debates on social issues, such as those presented today by artificial intelligence research and its industry applications. Experts in various matters relevant to citizen involvement on artificial intelligence can help implement these favourable conditions.

---

We have identified four (4) conditions required for the co-construction process: epistemic diversity, access to relevant information, moderation and iteration.

A. EPISTEMIC DIVERSITY
We must first ensure that the deliberating groups are as diverse as possible, in terms of social environment, gender, generation and ethnic origin. This diversity is not only inherent to the idea we have of an inclusive democracy, but is required to increase the epistemic quality of the debates. This simply means that every person brings a different perspective to the subject being debated, and thus enriches the discussion22.

B. ACCESS TO RELEVANT INFORMATION
We know, however, that epistemic diversity is insufficient and that if the participants have no skills or knowledge in the field being discussed, they cannot produce new knowledge, or follow the discussion. Collectively, then, they may increase individual errors. We must, therefore, prepare participants by providing relevant and quality information that is both accessible and reliable. An information session needs to be held prior to the deliberations.

C. MODERATION
Aside from having quality information, participants need to reason freely; that is, without being impeded by cognitive biases. We define cognitive bias as a distortion of rational thought by intuitive mechanisms. One of the most common and problematic biases in a deliberation is the confirmation bias: we tend to only accept opinions that confirm our own beliefs, and reject those that go against what we already believe. There are dozens of cognitive biases that can skew our logical train of thought.

But there are also biases that apply to the deliberation itself, such as the tendency to adopt more and more radical positions: if the group that is deliberating is initially distrustful of innovations in artificial intelligence, they will likely be entirely hostile towards them at the end of the deliberation. To avoid this type of knee-jerk reaction, we feel it is important to ensure epistemic diversity in the deliberating group and introduce a moderating body. This does not necessarily have to be personal input from a moderator. Although we do not object to individual moderation, we believe we can overcome deliberation biases through other means, such as introducing unexpected events in scenarios that sparked the discussions.

D. ITERATION
Ideally, we should be able to bring together the entire population to reflect on the responsible development of artificial intelligence. But the conditions described above cannot be applied to very large groups, let alone a society of several million people. It is important, then, to involve citizens in smaller groups and increase the number of meetings. This is the iteration phase of co-construction.

For these reasons, we prefer the structure of a co-construction workshop that brings together non-expert citizens, experts, stakeholders (associations, unions, professional representatives, businesses) as well as political representatives. These workshops are organized in different formats adapted to the deliberation spaces and satisfy the conditions for productive and robust citizen engagement. It is worth noting that the Declaration’s roll-out is complex, and relies on other types of consultations: online surveys, reports and expert roundtables. The Declaration is not simply a straightforward record of what was said in the co-construction workshops, but the result of multiple deliberations and reflections based on the co-construction workshops.

4.1.2 Experts and citizens

Why allow citizens to be heard on complex ethical and political questions that require a solid grasp of the technologies being discussed? Why not only consult the experts? There are many reasons, but the easiest is that AI affects everyone’s lives; therefore, it concerns everyone and everyone must have a say in the socially desirable goals of its development.

Even when we are not, strictly speaking, faced with a dilemma, public ethics issues cannot be resolved without making choices that favour certain moral interests over others, while still not neglecting them. This is the result of pluralistic values that define the moral and political context of modern democratic societies. It is possible to promote well-being by challenging the priority of consent: think of a medical app that could access personal data without our consent, but that would help treat serious diseases thanks to the data.

This type of ethical and social choice should be in the hands of all members of our democratic society, and not just a part, a minority, even if they are experts.

The role of experts’ role is not to solve the ethical dilemmas raised by artificial intelligence themselves, nor become legislators. What are the experts doing then? The experts involved in the co-construction process of the Montréal Declaration have no intention of thinking for citizens, nor suggesting a legal and ethical framework that citizens would merely rubberstamp. Expertise must be used to support citizen reflection on complex social and ethical AI issues.

---

Sometimes ethicists can appear preachy, possess the answers to difficult questions that the public itself is still asking, or seem to be able to solve tomorrow’s problems before they even happen. It is important to specify their role. Ethicists play three modest but crucial roles. They must:

> Ensure conditions exist that encourage citizen involvement
> Clarify ethical issues underlying the controversies around artificial intelligence
> Rationalize arguments being defended by participants by pointing out the arguments we know to be wrong or biased and explaining the reasons why they are wrong.

The role of ethicists is, therefore, to provide informed guidance. Experts in other research fields (computer sciences, health, safety, etc.) also play a role by providing participants with the most useful and reliable information on the subject of controversy (How does an algorithm that learns to make a diagnosis work? Can a doctor be replaced by a robot programmed to make a diagnosis? What protective measures can we introduce to thwart attempts to hack our medical data?, and so forth.)

However, we should acknowledge that experts themselves often show important cognitive biases. They can be too optimistic or pessimistic about new technologies they know well; they also tend to lend too much weight to their own opinion, especially when they believe they can predict the evolution of their field of research and social change. By involving them as citizens in the co-construction workshops, we reduce the biases associated with expertise, as well as the role of authority created by the discrepancy in knowledge between them and other participants.

The co-construction workshops are forums for participation that help guide the socially desirable development of AI and innovate through proposals that shake up traditional analysis frameworks. The vital contributions from citizen deliberations are then analyzed and expanded upon by work committees composed of experts from different fields (researchers, professionals). The work of expanding and drafting recommendations follows the guidelines defined by the deliberation and stays true to the proposals issued at the co-construction workshops.

4.2
THE CO-CONSTRUCTION WORKSHOP METHODOLOGY

The first version of the Montréal declaration on Responsible AI, presented November 3, 2017, during the Responsible AI Forum, is the foundation for the co-construction process. Schematically, after having defined the “what”? ("which desirable ethical principles should be gathered in a declaration on the ethics of artificial intelligence"), the new phase becomes a matter of predicting—with citizens and stakeholders—how ethical controversies surrounding AI could surface in the next few years (in the fields of health, law, smart cities, education and culture, the workplace, public services), then imagining how they could be solved (for example, with a device such as sector certification, a new stakeholder/mediator, a form or standard, a public policy or research program).

The goal of the co-construction process and its workshops is primarily to test the principles of the Montréal Declaration for Responsible AI using foresight scenarios. Ultimately, the process will help specify ethical issues per sector, and establish priority recommendations for the AI community.

---

More than ten co-construction workshops were held between February and October: three-hour world cafés in public libraries, and two important day-long co-construction workshops with various citizens, experts and stakeholders (at the SAT in Montréal, at the Musée de la civilisation in Québec City and the Centre Culturel Canadien in Paris²⁵).

Choosing to organize world cafés in public libraries is directly linked to how these spaces are being reinvented to provide public services in Quebec and Canada²⁶. By moving from a space that lends books to that of an inclusive “third space” that seeks to empower all citizens (e.g. with digital literacy services, citizen support, cultural mediation and discussion areas, lending tools and fab labs), public libraries will certainly play a key role in the responsible deployment of AI in Quebec and Canada.

The co-construction days were held in iconic spaces (Société des arts technologiques in Montréal, Musée de la civilisation in Québec) and focused primarily on uniting stakeholders and the very diverse disciplines that must work together to determine how AI should be responsibly deployed in society.

4.3
UNIQUENESS OF THE CO-CONSTRUCTION APPROACH

When compared with other AI ethics initiatives currently underway in the world, this co-construction process features three particularly original and innovative dimensions:

> Using foresight methods, with sector scenarios set in 2025 and the use of short stories to illustrate how ethical controversies on AI could surface or grow in the next few years (in the fields of health, law, smart cities, education and culture, the workplace). These 2025 scenarios, which present a variety of possible situations in a wide-open future, will be used to spark debate and

---

²⁵ We thank the Canadian Embassy in Paris for making the Paris workshop possible on October 9, 2018.

²⁶ Christophe Abrassart, Philippe Gauthier, Sébastien Proulx and Marie D. Martel, « Le design social : une sociologie des associations par le design? Le cas de deux démarches de codesign dans des projets de rénovation des bibliothèques de la Ville de Montréal », Lien Social et Politiques, 2015, n° 73, p. 117-138.
identify, specify or anticipate ethical issues per sector on AI deployment in years to come. These 2025 discussions can then help retroactively formulate concrete recommendations for 2018–2020, to help us work toward mutually beneficial goals.

Next, using participatory facilitation techniques in multidisciplinary “hybrid forums”\(^\text{27}\) that invite citizens and stakeholders to reflect on shared uncertainty and possible futures (to flesh out a scenario, come up with ways to respond to an ethical risk, suggest additions to the Declaration should an orphan issue arise, i.e. without a corresponding principle).

Lastly, paying attention to “paradigm biases” that have very powerful framing effects in how problems are viewed (e.g. tackling ethical issues of self-driving cars solely from the angle of the tramway dilemma, as MIT’s Moral Machine experience team suggests) and in the context of the “speed-distance” paradigm in transport design), in order ensure a plurality of issues and draw attention to still unknown or emerging situations in a rapidly changing context.

**Figure 4: Strategic forecasting: a three-step process**

**STRATEGIC FORECASTING:**
*a three-step process (Lab Ville Prospective)*

1. Collective exploration: what are the ethical AI issues per sector in 2025?
2. Reflexion: What are the recommendations for 2018-2020?
3. Collective action: collective experimentation and roll-out

This goal of the co-construction workshop is to create a learning path over the course of the workshops that turns into a versatile, user-friendly and reproducible discussion kit that could be published in “open source” at the end of the co-construction process.

Details of the world cafés and co-construction days can be found in the appendix.

4.4

WORLD CAFÉS OUTSIDE LIBRARIES

We would like to underscore the involvement of Pauline Noiseau and Xavier Boileau, two philosophy students at Université de Montréal, who have organized many world cafés in non-library spaces, and who used a format that encourages more organic kinds of discussions on AI issues. Moderators used very short scenarios, and hosted 2-hour sessions. These sessions sparked meaningful deliberations among citizens who wanted nothing more than to be involved in public debates, but who were rarely asked to participate in them. That’s how a world café at the Maison d’Haïti, on April 25, 2018, allowed high school youth and retirees from the Saint-Michel neighbourhood in Montréal-Nord to trade opinions on AI issues. From an AI scenario on household connected objects (a smart refrigerator), this session sparked original ideas on cooking as a relational human activity, raising issues of authenticity, affection (“a touch of love”) and social ability, issues that had not come up in other consultations based on the same scenario.

4.5

PORTRAIT OF PARTICIPANTS

By recruiting citizens, experts and professionals from different fields of work, we had access to a diverse pool of participants in the co-construction workshops. We were also able to contact numerous stakeholders involved in AI development through university faculties as well as inter-university research centres and their networks.

The websites and social media of our different partners played an important role in soliciting the public, although local recruitment efforts by each library involved in the project proved to be the most efficient.

Of note: nearly as many men as women took part in all workshops. The majority of participants had a post-secondary education and fell into the 19-34 age group.
Figure 7: Distribution of participants in world cafés and co-construction days per level of education

**EDUCATION**

- No certificate, diploma or degree: 3
- High school diploma or equivalent: 2
- Postgraduate studies: 6
- College degree: 14
- University certificate below bachelor degree: 9
- Bachelor degree: 56
- University certificate above bachelor degree: 87
- Medical diploma: 2
- Doctorate acquired: 39
Figure 8: Distribution of participants in world cafés and co-construction days by area of activity

FIELDS OF ACTIVITY
34% of respondents indicated more than one field of activity

- Public administration: 17
- Arts, concerts and leisure: 16
- Other: 49
- Retail trade: 3
- Energy and resources: 3
- Teaching: 36
- Finance and insurance: 12
- Company and entreprise management: 6
- Hospitality: 1
- Information and culture: 15
- Research (industrial or university): 46
- Professional services, scientific and technical: 28
- Healthcare, biotechnology and social assistance: 17
- IT: 63
- Transportation and warehousing: 3
5. WORKSHOP DELIBERATIONS: EXAMPLES FROM SMART CITIES AND THE WORKPLACE

5.1 THE DELIBERATION PROCESS

How did the discussions and deliberations in the co-construction workshops unfold? What kinds of reactions did they elicit? What were the main points of discussion that led to recommendations for an AI framework? This section of the document includes highlights from the deliberations between participants, where each person took care to specify the reasons, principles and values justifying their position on the prospective scenario suggested as a starting point, whether it was to agree, disagree, nuance or question something. In a word, to do what pragmatic sociology has defined as justification.

To illustrate this work, the deliberations of two teams representing two of the five sectors discussed in the co-construction were selected:

- A table of citizens that discussed the self-driving car (smart city sector)
- A table of researchers and experts dealing with the impact of AI on jobs in businesses (workplace sector).

To formulate these recommendations, each team followed three steps where ideas were generated, then deliberated:

- **First step**: identifying ethical and social issues per sector in 2025 (by cross-referencing the general principles of the Montréal Declaration with the 2025 user situations described in debate scenarios): determining individual issues (on Post-its), which were then expanded upon in a collective discussion where three priorities were identified.

- **Second step**: proposing recommendations to be implemented in 2018-2020 to prepare for the responsible deployment of AI in Quebec: from formulating recommendations to choosing a few newspaper headlines.

- **Third step**: storytelling for the deployment of the first recommendation in 2020 (the newspaper headline) to consider the “time for collective action” with its organizational constraints: from coming up with ideas to synthesizing them in an orderly fashion within a narrative.

It is important to note that between these steps and micro-steps of the deliberations, the “nature” of the ideas generated varies: some are individual intuitions (when, at the start of the exercise, participants write down many sector issues on Post-its), others stem from a collective discussion (where each person justifies their point of view) and yet others are the result of a hierarchy determined by the group (when selecting three key issues to write on the summary poster).

We note three properties of deliberative mechanisms in these foresight workshops, as discussed by Blondiaux and Sintomer in their article *L’impératif délibératif* (Politix, 2002, pp. 25-26): allow new solutions to be imagined in an uncertain world; let generalities arise and aim for consensus or “deliberative disagreements” in a society characterized by value pluralism; and finally, provide a factual and normative source of legitimacy by including everyone in these deliberations.

---

5.1.1 Smart city sector: self-driving cars (SDC) and sharing the road equitably

**Summary of the initial 2025 scenario.** In 2025, the first SDCs are circulating in Montréal and controversy arises over sharing the road and public spaces. Some lanes are now reserved for SDCs and protected by barriers, so that they can drive at a moderate, but fluid speeds (50 km/h) without the risk of accident. SDCs can also drive elsewhere, but at very slow speeds (25 km/h). Proponents of active mobility (walking, biking) disrupt these protected lanes, knowing that SDC algorithms are set to "altruistic" mode to protect people outside of them.

The goal of this scenario was to open a discussion on the ethical issues of SDCs with a situation that recreated the density and complexity of a city: low speeds and different speeds, fluidity as a priority criteria for speed, protection barriers for safety, the road as a shared space for competing uses.

The deliberations presented are the result of a three-hour roundtable in a Montréal public library, with eight citizens keen on new technologies, whose families embrace active transportation (walking, biking). From this 2025 scenario, the discussion led to an initiative presented as a headline in the March 13, 2020, edition of the Responsible AI Gazette: "First autonomous mobility literacy workshop held."

What were the deliberations that led to this unique proposal? What were the defining moments? How did the ideas evolve at each step? We present and comment on certain highlights of deliberations by this team.

**Highlights from first deliberations: FRAMING ETHICAL ISSUES IN 2025**

A number of questions about different principles of the Montréal Declaration were written on Post-its and submitted by participants:

**Autonomy:**

"Will humans become too dependent when it comes to getting around?", "Will freedom of movement be impeded by AI?", "We’re entrusting AI and interconnected systems with making a lot of micro-decisions to, at the expense of humans."

**Well-being:**

"A lot less room for spontaneity with SDCs", "How will neighbourhoods and SDC roads be developed?", "Will transportation data influence the urbanization of cities?"

**Democracy and justice:**

"How will building roads in working-class neighbourhoods and affluent neighbourhoods differ?", "Will only those who are well-located get to take advantage of fluid traffic?"

**Privacy:**

"Will we be able to track everyone’s movements?"

Responsibility: "Who will be held responsible for an accident?" Security: "Can fleets of vehicles be hacked?" This last principle came from participants, in addition to those found in the declaration.
Many in-depth discussions then took place, with participants bouncing off initial ideas and generating new ones on spontaneity and freedom to travel, the safety of personal data being managed by a central organization, algorithm settings and their potential for manipulation.

After a nearly 45-minute-long discussion, the participants used coloured stickers to select ethical issues for 2025, which were grouped by priority. Votes were cast via coloured Post-its on the wall and ideas associated with four principles of the Montréal Declaration were discussed, two of which were regrouped: safety, justice, and well-being and autonomy.

Table 1: Smart City, Highlights from first deliberations: framing ethical issues in 2025

<table>
<thead>
<tr>
<th>2025 Ethical Issues</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Ease of hacking centralized system. Dilemma: collective fluidity - system vulnerability</td>
<td>Risk of social exclusion Settings classification by social class (e.g. trip through poor neighbourhood - VIP settings)</td>
<td>Loss of spontaneity when travelling, loss of independence and freedom of movement, and geo-localization.</td>
</tr>
<tr>
<td>Associated Principles</td>
<td>Security</td>
<td>Justice</td>
<td>Well-being and autonomy</td>
</tr>
</tbody>
</table>

The priority issues selected by the team are unique: although issues of security, responsibility and privacy are often raised in studies and debates on SDCs, those of justice, well-being and autonomy are not discussed as often.
Highlights from second deliberations:
RECOMMENDATIONS FOR AN AI FRAMEWORK IN 2018-2020

In response to these issues, the team chose to pursue discussions by continuing to reflect on the four associated principles. Participants proposed a number of recommendations for an AI framework. We present three (out of six) here, which allow you to see how an idea evolved into a newspaper headline.

Table 2: Smart City, Highlights from second deliberations: recommendations for an AI framework in 2018-2020

<table>
<thead>
<tr>
<th>Framework recommendations for 2018-2020</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training for collective vigilance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g. driver’s licence)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An all-party committee that manages incidents, infractions and other issues in democratic fashion; it must be a decision-making committee</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluating urban development plan during the transition period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tool Categories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New institutional stakeholder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participative planning process</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These recommendations, which demonstrate true institutional creativity (beyond the very broad examples of tools provided in the participant booklet), dovetail with issues identified in the previous step, but also generated more robust ideas (they are not merely deductions from tools adapted from an identified ethical case). The idea of vigilance training and participating in collective decision-making (through an all party committee and open planning) led to recommendations for capacity-building and local forms of democracy.
Highlights from third deliberations: WRITING A HEADLINE AND LEAD FOR A 2020 NEWSPAPER

These measures were then storyboarded on the poster. The headline of the March 13, 2020, edition of the Responsible AI Gazette designed by the team read as follows:

"FIRST AUTONOMOUS MOBILITY LITERACY WORKSHOP HELD"

“The Quebec public library network has introduced a training program on the use of self-driving cars. On the curriculum: collective vigilance; the code of ethics; how to get involved in the city’s decision-making process; roadsharing between pedestrians, bicycles, SDCs, trucks; understanding the rules; trial sessions; the issue of algorithm settings."

This newspaper headline, which was drafted following a discussion among the participants, once again demonstrates how ideas evolve. The principle of a workshop on “autonomous mobility literacy” allows new meaning to be created by integrating various recommendations proposed in the previous step, thus widening the scope to autonomous mobility and not merely SDCs (thereby including the possibility of autonomous multimodal transportation). This headline also presents a collective action measure with a progress target (training and capabilities of citizens, the possibility of joining city decision-making committees on SDC deployment) and organization (roll-out in public libraries across Quebec, which are now transforming into cultural “third spaces” for citizens).

The result of this roundtable is particularly interesting because it allows us to consider the ethical question of self-driving vehicles from the perspective of autonomy and social justice in the city, and not strictly responsibility in the event of an accident, as MIT’s Moral Machine initiative does, for example, with the moral dilemma of the tramway.\footnote{MIT site: http://moralmachine.mit.edu}
5.1.2 Workplace sector: Socially responsible restructuring?

**Summary of the initial 2025 scenario.**
In 2025, many businesses use AI in their management tools. Such is the case for an eco-friendly logistics company that must make a massive investment in AI and robotics to remain competitive. Parcel sorting, routing, administrative follow-up, calculating the carbon footprint of the trips, self-driving electric trucks: in total, up to one third of the company’s positions could be cut. The company, which is very socially involved, wants to proceed with restructuring in a socially responsible manner, for instance by creating a data processing co-op to rehire as many salaried employees as possible, independently from the big corporations. Will it be able to do so in time?

The goal of this scenario was to spark a discussion on the ethical and social challenges related to change processes caused by AI that thousands of SMEs and big businesses in Quebec will face between 2020-2030.

The deliberations presented in this section were held during a day-long roundtable in Montréal that brought together nearly 10 researchers and experts working on workplace mutations, social participation and social responsibility of businesses and unions. A citizen that had previously attended a workshop in a public library also took part in this roundtable.

Working from this 2025 scenario, this team came up with an initiative that made the headline of the February 18, 2020, Responsible AI Gazette: **“First measures of the joint interdepartmental committee on responsible digital transition.”** As in the previous case on self-driving cars, how did the deliberations lead to this unique proposal? What were the defining moments? How did the ideas evolve at each step? We present and comment on highlights of the team deliberations.
Highlights from first deliberations:
FRAMING ETHICAL ISSUES IN 2025

Participants wrote numerous ideas on Post-its in the first half of the morning workshop. Here are a few of them and a sampling of some ideas taken from Post-its, which were grouped according to principles of the Montréal Declaration.

Some ideas were associated with different principles of the Montréal Declaration:

WELL-BEING:

“What should we promote? The company or society?”, “Adopting different perspectives on well-being: individual (employee), social and collective development, economic development (SME)”, “What does performance measurement look like when robots or co-bots never get tired, unlike humans?”, “What are the possible positive aspects: professional support, e.g. in medicine, less drudgery in certain positions”, “What are the new forms of work and protection with work/leisure?”

AUTONOMY:

“What sort of career and life paths? Can you choose not to change careers because of AI? What are the consequences?”, “collective autonomy: for collectively and critically anticipating discussion on the urgency of adapting”

RESPONSIBILITY:

“Who is held responsible for these changes?”, “Is the social and ethical responsibility of the transition individual—each company—or collective—society, the government?”, “Where does for the transition funding come from?”; “How can we align cost-effective measures and responsibility in an emergency situation?”

KNOWLEDGE:

“What will collaboration between humans and robots look like? Workload, health and safety, training, acceptability, cybersecurity,” “How is data collected in settings where this type of work is mainly carried out by private corporations (GAFAM)?”, “How can we prevent people from getting pigeonholed in classes?”, “What are the possibilities of data being shared?”, “What is the impact on the educational system?”

THE JUSTICE PRINCIPLE:

“What independence exists when power is concentrated among GAFAM?”, “How will the benefits of AI be distributed among society?”, “Will productivity gains created by AI and industry 4.0 be sufficient to fund social change if companies engage in tax evasion?”, “What fairness exists when sharing and coding an employee’s implicit knowledge so that it can be transform into data or feed automation?”, “Do we have a choice, as employees, not to reveal this information?”, “What is the criteria for choosing those who are replaced and those who are trained?”, “Will social protection of tomorrow be accessible?”, “What access to rights, such as right of association, in this new, reorganized workplace?”

DEMOCRACY:

“Is job insecurity inevitable when transition can be anticipated?”, “the politicized short-term vision over a long-term vision”, “obscuring decision-making processes”, “risks of bias in algorithm training sets”, “the need for a democratic debate”
We should note that the classification of the Montréal Declaration on responsible AI principles was useful in providing discussion benchmarks, and that the participants even came up with unique issues on certain principles: the need to address well-being and responsibility for the transition from different points of view (individual and collective); the relationship with social time, with collective anticipation and the opaque language of urgency opposed to one another, as a condition of our collective autonomy and exercising our democracy (lack of time preventing well-informed democratic work); a strong need for justice in the social redistribution of AI benefits, namely in terms of equity accompanying codification, and therefore possible automation, of employee skill sets.

After a good hour of discussion, participants used coloured stickers to determine groups of 2025 ethical issues they deemed priorities. With votes spread out fairly evenly on various issues, all deemed equally important by the group, three priorities for the poster were identified after the ideas discussed in the first half of the workshop were synthesized (see table below).

Table 3: Workplace, Highlights from first deliberations: framing ethical issues in 2025

<table>
<thead>
<tr>
<th>2025 Ethical Issues</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massive concentration of power (see GAFAM) that prevents:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Equitable sharing of AI benefits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Arrival of new stakeholders (new co-op type business models)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Inequities to be minimized (literacy)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technological determinism, inevitability (&quot;Black box society&quot;) and urgency: instead of taking the time to have an informed, participative, democratic debate on new social risks, social development models, performance measures, work experience.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defining the common good and the kind of collective responsibility in the digital transition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For example: Which stakeholders?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The company alone?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The State?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unions?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The educational system?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Associated Principles</strong></td>
<td>Justice and independence</td>
<td>Democracy, knowledge and collective autonomy</td>
<td>Well-being and responsibility</td>
</tr>
</tbody>
</table>
Second deliberations: RECOMMENDATIONS FOR AN AI FRAMEWORK IN 2018-2020

To respond to these issues, the team resumed talks in the afternoon by leading another roundtable during which participants drafted recommendations for an AI framework. This then led to numerous recommendations that were discussed one by one as a group. The table below presents an excerpt (six out of the more than 10 proposals that were formulated by the group), in order to follow the evolution of an idea up to a newspaper headline.

Table 4: Workplace, Second deliberations: recommendations for an AI framework in 2018-2020

<table>
<thead>
<tr>
<th>2018-2020 framework recommendations</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reinforce digital literacy for all.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With a skill set reference software for public libraries, schools, and the workplace. By tackling the question of illiteracy and “non-use” by citizens.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint permanent interdepartmental committee on AI, executive next to PM.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interface between themes of economy, employment, education and culture. (see digital strategy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital AI insurance funds to make way for training and transition.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of measure: a 50-week Parental Insurance Plan which can also generate a minimum income to prevent job insecurity.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentives on new business models for data processing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example: Co-op model that prevent self-employed workers processing data from becoming isolated and guarantees collective autonomy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investing in responsible AI for the common good.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRI (Socially responsible investment) model. Investments from the State, individuals, in synergy with the worker’s fund.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accelerated process to update and create professional programs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With cégeps, universities, departments, professional orders impacted by AI (e.g. law, healthcare).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tool Categories</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New institutional stakeholder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New insurance mechanism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funding device</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planning process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As in the previous case of self-driving cars, these recommendations, which demonstrate true institutional creativity (beyond the very broad examples of tools provided in the participant booklet), dovetail with issues identified in the previous step, but also generate more robust ideas. If digital literacy is indeed a goal in the policy’s agenda (e.g. Stratégie numérique du Québec), the need to expand it was highlighted. Other recommended measures are innovative and call for the creation of new public, all-party or collective measures to ensure the true autonomy of Quebec society on AI issues in the workplace. To that end, the group chose collective responsibility towards AI in its transition into society.
These measures were then storyboarded for the poster. The headline drafted by the team for the February 18, 2020, Responsible AI Gazette reads as follows:

**FIRST MEASURES OF THE JOINT INTERDEPARTMENTAL COMMITTEE ON RESPONSIBLE DIGITAL TRANSITION**

The new committee, created on March 14, 2018, after the co-construction workshop for the Montréal Declaration for Responsible AI, quickly got to work and developed a coherent strategy integrating all stakeholders. In early 2020, the committee was proud to announce the launch of four (4) programs:

1. A new digital insurance fund worth $2 billion (funded by productivity gains attributed to AI).
2. An agreement with all cégeps and universities to accelerate the renewal of training programs.
3. A support program to create self-employed worker co-operatives (against job insecurity).
4. A five-year literacy fund worth $10 billion based on a new skill set inventory.

This newspaper headline, which was drafted following discussion among participants, once again helps ideas evolve. The joint interdepartmental committee on responsible digital transition was entirely new. This new institutional stakeholder, born of a reflection on a 2025 scenario that dealt with the impact of AI on the Quebec workplace, could represent a new common step for many public policies that successfully address digital transition and digital literacy, but do not tackle the question of AI’s social impact: the *Stratégie numérique du Québec* du ministère de l’Économie, de la Science et de l’Innovation (MESI), the *Stratégie nationale sur la main-d’œuvre 2018-2023* du ministère du Travail, de l’Emploi et de la Solidarité sociale (MTESS), the *Plan stratégique 2017-2022* du ministère de l’Éducation et de l’Enseignement supérieur (MEES).

This new stakeholder, possibly the result of a cross collaboration between the Commission des partenaires du marché du travail (CPMT), the Comité consultatif sur le numérique and the Commission mixte de l’enseignement supérieur, would specifically anticipate workplace transformations and new training and adaptation issues caused by the deployment of AI in Quebec’s public and private organizations.
6. PARTICIPANTS IN THE CO-CONSTRUCTION AND WORKING GROUPS

Citizens, professionals and experts who took part in the workshops, in Québec and Paris, and agreed to let us publish their names:

Sihem Neila Abtroon
Sébastien Adam
Béatrice Alain
Hassane Alami
Rana Alvabi
Alejandro Arreola-Alvarado
Gabriel Arruda
Jean-Claude Asssaker
Barthélémy Aucourt
Naomi Ayotte
Manon Babine
Maryluisa Barillas
Philippe Beauchemin
Stéphane Beaulieu
François Beauregard
Claude Bédard
Sylvain Bédard
Abdelkader Bekhti
Halim Benzaïd
Vincent Bergeron
Alexandre Berkesse
Karl Bherer
Emmanuel Bloch
Marise Bonenfant
Serge Bouchard
Caroline Boudreault
Lyne Bourbonnais
Véronique Boutier
Morgane Bravo
Robert Bruno
Beatrice Cassar
Ofelia Castaneda
Chantal Caux
Christian Chabot
Michel Chabot
Karine Charbonneau
François Charbonnier
Anne Chartier
Philippe Chartier
Guillaume Chicoisne
Pierre Choffet
Dominic Cliche
Lilien Colombino
Cristina Cotargasanu
François Côté
Jacques Coulombe
Lise Couturier
Alexis Cuglietta
Christian Cyr
Yvonne Da Silveira
Geneviève Dagneau
Hélène David
François-Michel De Rainville
David Décair-Hétu
Guillaume Déraps
Yves B. Desfossés
Michel Desy
Marc-Antoine Dilhac
Maxime Duban
Jean-Yves Dubé
Geneviève Dubois-Flynn
Mathieu Dubreuil-Cousineau
Geneviève Dufour
Arnaud Duhoux
Annie Dulude
Laurence Dumont
Mathieu Dumouchel
Benoit Dupont
Nicolas Dupras
Diane Duquette
Irina Entin
Julian Falardeau
Jacqueline Forien
Simon Frappier
Benoit Gagnon
Marie-Pierre Gagnon
Marina Gallet
Hortense Gallois
Sébastien Gambs
Véronique Gareau-Chiasson
Mathieu Gauthier-Pilote
Sylvie Gélinas
Thomas George
Gueno Gianni
Jean-François Gignac
Martin Gibert
Patricia Gingras
Béatrice Godard
Co-construction team – In Québec and Paris

Simon Beaudoin-Gagnon, Maison des étudiants canadiens
Alexandre Beaudoin-Peña, Université de Montréal
Bhavish Beejan, Université Laval
Liam Bekirsky, Maison des étudiants canadiens
Karl Bherer, Université Laval
Alexis Bibeau, Université Laval
Pierre-Antoine Boutin-Panneton, Université Laval
Katie Charpentier-Bourque, Maison des étudiants canadiens
Arnaud Brubacher-Chouinard, Maison des étudiants canadiens
Dominic Cliche, Université Laval
Valentine Crosset, Université de Montréal
Rosemarie Desmarais, Maison des étudiants canadiens
Eve Gaumond, Université Laval
Martin Gibert, IVADO and Centre de recherche en éthique éthicien
Emilie Guiraud, Université Laval
Haykuhi Gutrez, Maison des étudiants canadien
Hubert Hamel-Lapointe, Université de Montréal
Audrey Houle, Université Laval
Samira Illourman, Maison des étudiants canadiens
Nico Julien, Université Laval
Henri Lajeunesse, Université Laval
Lauriane Long-Raymond, Maison des étudiants canadiens
Guillaume Macaux, Université Laval
Vincent Mai, Université de Montréal
Mariève Mauger-Lavigne, Université de Montréal
Christophe Mondin, CIRANO
Orly Nahmias, citizen
Pauline Noiseau, Université de Montréal
Judith Paquet, Université Laval
Pierre-Luc Plante, Université Laval
Léa Ricard, Université de Montréal
Lynda Robitaille, Centre de recherche en données massives, Université Laval
Jason Stanley, Université de Montréal
Yanis Taleb, Université de Montréal
Clémence Varin, Université Laval
Nathalie Voarino, Université de Montréal
Camille Vézy, Université de Montréal
Alessia Zarzani, Université de Montréal
Consulted experts

Sylvain Bédard, Patient Coordinator at Centre of Excellence on Partnership with Patients and the Public (CEPPP)
Louise Béliveau, Vice-Rector of Student and Academic Affairs, Université de Montréal
Guillaume Chicoisne, Scientific Programs Director, IVADO
David Décary-Hétu, Assistant Professor, School of Criminology, Université de Montréal; Regular researcher at Centre international de criminologie comparée
Pierre-Luc Déziel, Professor, Faculty of Law, Université Laval
Thierry Karsenti, Full Professor, Faculty of Education, Université de Montréal; Chairholder of the Canada Research Chair on information and communication technologies in education
Jihane Lamouri, Diversity Coordinator, IVADO
Lyse Langlois, Full Professor and Vice-Dean of the Faculty of Social Science; Director of the Institut d’éthique appliquée (IDÉA); Researcher Interuniversity Research Centre on Globalization and Work (CRIMT)
François Laviolette, Full Professor, Computer Science, Université Laval; Director of the Centre de recherche en données massives (CRDM)
Marie Martel, Professor in École de bibliothéconomie et des sciences de l’information, Université de Montréal
Nicolas Merveille, Professor in École des sciences de la gestion (ESG), Université du Québec à Montréal (UQAM); Co-chairholder of the International Life Cycle Chair, ESG UQAM and École Polytechnique; Supervisor for the City of Montréal’s Internet of Things Ethics and Social Acceptability project
Gregor Murray, Director of Industrial Relations and Interuniversity Research Centre on Globalization and Work (CRIMT)
Catherine Régis, Associate professor, Faculty of Law, Université de Montréal; Chairholder, Canada Research Chair in Collaborative Culture in Health Law and Policy; Regular researcher, Centre de recherche en droit public (CRDP)
Nicolas Roby, Scientific Coordinator of Industrial Relations and Interuniversity Research Centre on Globalization and Work (CRIMT)
Frank Scherrer, Full Professor, Urbanisme, Université de Montréal; Director at the École d’urbanisme et d’architecture de paysage, Université de Montréal; Academic Director of the EDDEC Institute, an organization promoting the environment, sustainable development, and circular economies
Marie-Odette St-Hilaire, Architecte de solutions TI, Science de données, Service des technologies de l’information, Ville de Montréal

Management team

Isabelle Bayard, Vice-Rector assistant of Reasearch, Discovery, Creation and Innovation, Université de Montréal
Joliane Grandmont-Benoit, Project Coordinator, Vice-rectorate of Student and Academic Affairs, Université de Montréal
Anne-Marie Savoie, Adviser, Vice-Rectorate of Reasearch, Discovery, Creation and Innovation, Université de Montréal
Research and analysis team

Valentine Crosset, Doctoral student in criminology, Université de Montréal
Jean-François Gagné, Programs Adviser, Political Science, Université de Montréal
Vincent Mai, Doctoral student in robotics, Université de Montréal
Mario Ionut Marosan, Master in political philosophy, Université de Montréal
Marie Martel, Professor in École de bibliothéconomie et des sciences de l’information, Université de Montréal
Loubna Mekki-Berrada, Doctoral student in neuropsychology, Université de Montréal
Christophe Mondin, Research professional for CIRANO
Camille Vézy, Doctoral student in communication, Université de Montréal
Nathalie Voarino, Scientific Coordinator, PhD Candidate in Bioethics of Université de Montréal
Alessia Zarzani, Doctoral student in Planning, Université de Montréal

Coordination team in Paris

Jacques-Henri Gagnon, Head, Communication, Youth and Academic Relations, Embassy of Canada to France
Hanane Hadjiloum, Communications Officer, Maison des étudiants canadiens
Christine Métayer, Director of the Maison des étudiants canadiens
Clément Thiébault, Trade Commissioner, France (ICT), Embassy of Canada to France

Partners who contributed to the fall co-construction workshops

Students from the Comité intersectoriel étudiant (CIÉ) from the Québec Research Funds, participants in the Journées de la relève en recherche held by ACFAS
Professional members from the Coalition for the Diversity of Cultural Expressions (CDCE–Canada)
Elected officials and employees from trade unions that participated in the AI discussion forum organized by the Syndicat de la fonction publique et parapublique du Québec (SFPQ)
ANNEX 1 – CO-CONSTRUCTION WORKSHOPS: HOW THEY WORK

World cafés

World cafés are three-hour-long meetings in public libraries. These meetings are inclusive, open to all citizens, and take place in a friendly atmosphere. These meetings will be based on the World Café model.

The world café provides an enjoyable forum for conversation and seeks to encourage constructive dialogue and the exchange of ideas. The goal is to recreate a café ambiance where participants debate a question in small groups. At regular intervals, participants change tables. One host stays at the table and sums up the previous conversation for the people who have just arrived. The ongoing conversations are therefore "pollinated" by ideas from previous conversations. At the end of the process, the main ideas are summed up during a plenary session, and possible follow-ups are submitted for discussion.

This world café technique was adapted and enhanced by many elements:

> Presenting the Montréal Declaration and the social and ethical issues of AI
> Reading sector foresight scenarios set in 2025 to kick off discussions
> Using a poster to document the discussions
> Handing out a participant workbook presenting the principles of the Montréal Declaration for Responsible AI, a lexicon and an sample classification of possible recommendations.

Here is what a typical world café looks like:

<table>
<thead>
<tr>
<th>Steps</th>
<th>Time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welcome</td>
<td>1 p.m. to 1:30 p.m.</td>
<td>Coffee and snacks</td>
</tr>
<tr>
<td>Discovering AI and its ethical and social implications</td>
<td>1:30 p.m. to 2 p.m.</td>
<td>Educational Introduction: introduction to the ethical and social implications of artificial intelligence (Montréal Declaration), presentation of scenarios set in 2025 and the activity.</td>
</tr>
</tbody>
</table>
| World café | 2 p.m. to 4 p.m. | > Four thematic islands (on AI in health, justice, education, smart cities and the workplace) are hosted by a facilitator. Each island hosts a small group of participants (6 to 10) for two 50-minute discussions about an AI scenario set in 2025.  
> participants are invited to imagine the “front page of a 2020 newspaper” (headline and first paragraph) discussing an important initiative in Quebec for the responsible deployment of AI. |
| Summary in plenary session | 4 p.m. to 4:30 p.m. | Summary of discussions during a plenary session. The facilitators sum up the posters from each thematic island, followed by a group discussion. |

Tableau 5: Déroulement type des cafés citoyens

---

30 Definition from the Institut du nouveau monde (INM)
Co-construction workshops

These one-day meetings brought together citizens, stakeholders and experts seeking to explore sector issues and develop recommendations. They are based on the foresight co-design model, developed at the University of Montréal’s Lab Ville Prospective.

The workshops are based on the foresight co-design model that combines design, participation and forecasting: imagining scenarios and unknown prototypes as conversation starters, opening up cognitive possibilities and pathways for exploration (the design dimension); using collective participation techniques that bring together stakeholders from diverse backgrounds, citizens and organizations as experts (for the collective aspect of the “co”); finally, the foresight approach which consists of projecting oneself into a possible future 10 or 20 years from now to make an imaginary detour and then work back from there to develop innovative paths that link the present to the most desirable futures. Michel De Certeau, in his work La culture au pluriel (1993, p. 223) highlights otherness in foresight, saying, "the future engages the present in the mode of otherness". And Georges Amar, in an article on conceptive foresight (in Futuribles, 2015, p. 21) insists on the importance of creating a narrative around the unknown to build an open future: “We prefer inefficient known properties to the promising unknown. The purpose of foresight is to work on the unknown, to give words, concepts, language on it. So that while it remains unknown, it becomes more accessible, leads to reflection, and action.’

Here is what a typical co-construction workshop looks like:

Table 6: What a typical co-construction workshop looks like

<table>
<thead>
<tr>
<th>Steps</th>
<th>Time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welcome</td>
<td>8:30 to 9 a.m.</td>
<td>Coffee and pastries</td>
</tr>
<tr>
<td>Introduction and AI Discovery</td>
<td>9 a.m. to 10 a.m.</td>
<td><strong>Introductions:</strong> principles of artificial intelligence, ethical issues surrounding AI (Montréal Declaration) and foresight scenarios.</td>
</tr>
<tr>
<td>Foresight team</td>
<td>10 am to 11:30 am</td>
<td><strong>Team foresight:</strong> starting with a trigger event and the Montréal Declaration principles, frame the ethical and social issues raised by the 2025 scenario and explore how an ethical controversy could arise or grow.</td>
</tr>
<tr>
<td></td>
<td>11:30 am to 12:30 pm</td>
<td><strong>Plenary:</strong> Plenary presentation of 2025 ethical and social issues raised, and discussions with entire group.</td>
</tr>
<tr>
<td>Lunch on site</td>
<td>12:30 pm to 1 pm</td>
<td>Lunch</td>
</tr>
<tr>
<td>Developing recommendations</td>
<td>1:30 pm to 2:45 pm</td>
<td><strong>Developing recommendations</strong> Work in teams: using the 2025 ethical issues identified in the morning, develop recommendations (rules, sectoral codes, labels, public policies, research programs, etc.) to implement in 2018–2020 in Quebec.</td>
</tr>
<tr>
<td></td>
<td>3 pm to 4 pm</td>
<td><strong>Plenary team presentations</strong> and group discussion</td>
</tr>
<tr>
<td>Conclusion and follow-up</td>
<td>4 p.m. to 4:30 p.m.</td>
<td>Review and observations surrounding the day</td>
</tr>
</tbody>
</table>
This annex presents a summary of all the AI scenarios used in this first co-construction phase, and five complete scenarios. Set in 2025, in Quebec, they were the starting point for debates and deliberations on the ethical questions raised by artificial intelligence. The year 2025 was selected as it was in the near future, at the heart of the decade 2020–2030 which should see intensive deployment of artificial intelligence in society.
1. Scenarios per theme

Eighteen scenarios were debated from February to May 2018. The table below presents a summary of these scenarios.

<table>
<thead>
<tr>
<th>Theme</th>
<th>2025 AI scenario</th>
<th>Summary of AI scenario in 2025 in Quebec</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Predictive Health</td>
<td>Healthy digital twins</td>
<td>Olivier learns that one of his 126 digital twins has been diagnosed for depression. Should he go see a professional?</td>
</tr>
<tr>
<td></td>
<td>Discriminating Health Insurance</td>
<td>Olivier’s insurance company asks him to change his lifestyle, based on his personal data. Can he refuse without suffering any consequences?</td>
</tr>
<tr>
<td></td>
<td>Vigilo, a house robot for the elderly</td>
<td>Soline is 80 years old and lives at home with Vigilo, her robot companion. Her robot regularly reports predictive diagnoses on Soline’s health to her family. Does she want everything revealed?</td>
</tr>
<tr>
<td></td>
<td>A therapeutic decision at the hospital</td>
<td>An experienced doctor and a medical recognition algorithm do not quite agree on a diagnosis.</td>
</tr>
<tr>
<td></td>
<td>Self-driving cars (setting the algorithm and sharing the road)</td>
<td>To ensure its zero-accident policy, the city has established safety barriers on roads where self-driving vehicles can go “fast” (50 km/h). A controversy on sharing the road ensues.</td>
</tr>
<tr>
<td></td>
<td>Self-driving cars (restricted use)</td>
<td>Self-driving cars have become a rideshare service for citizens. Priority access criteria is managed by AI to maximize the city’s predictive economic growth.</td>
</tr>
<tr>
<td>2. Smart City</td>
<td>A connected fridge that wants what’s best for you (nudges)</td>
<td>A family purchased a smart fridge with a “nudge” program to encourage healthy eating and reduce risks of disease. How will the gains from this system be divided between the insurance company and the family?</td>
</tr>
<tr>
<td></td>
<td>A social rating based on a carbon footprint</td>
<td>A family’s consumption is defined and tracked to prevent a negative impact on the environment.</td>
</tr>
<tr>
<td></td>
<td>A smart toy that’s not all that loyal!</td>
<td>How far does a smart toy’s loyalty to a child go? Is it the same as a friend’s?</td>
</tr>
<tr>
<td>Theme</td>
<td>2025 AI scenario</td>
<td>Summary of AI scenario in 2025 in Quebec</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>3. Predictive education</td>
<td>AlterEgo, AI that assists learning at school</td>
<td>AI helps students learn more efficiently, thanks to personalized homework and exercises. Does the teacher still have complete professional autonomy?</td>
</tr>
<tr>
<td></td>
<td>AlterEgo2, AI School Guidance Assistant</td>
<td>AI guides students towards careers where the odds of succeeding are very strong. Based on their history of school data, will the choice really reflect the student’s wishes?</td>
</tr>
<tr>
<td></td>
<td>Nao, AI that helps prepare conferences</td>
<td>AI helps a lecturer develop his presentation and update it throughout the lecture, according to the reactions of his students.</td>
</tr>
<tr>
<td>4. Police and predictive justice</td>
<td>A preventive arrest in a public space</td>
<td>Cross-referencing Alexandre’s personal data has recently flagged him as an individual who is potentially at risk. After acting strangely in a public space, he is arrested preventively.</td>
</tr>
<tr>
<td></td>
<td>A parole decision</td>
<td>A judge makes the decision to order probation for a detainee, against the algorithm’s recommendation. The algorithm anticipates likely recidivism, but without taking into consideration a new reinsertion program (without any data history).</td>
</tr>
<tr>
<td>5. Workplace</td>
<td>AI to improve workplace atmosphere</td>
<td>A company’s human resources department uses AI with data mining to evaluate the behavioural style of their employees and help them to cultivate a “good workplace atmosphere”.</td>
</tr>
<tr>
<td></td>
<td>Recruitment AI as a compulsory step to employment</td>
<td>All candidates for a position will be recruited according to a video analyzed by AI, in order to eliminate any bias, favourable or not. Is recruitment neutrality real, and is it desirable?</td>
</tr>
<tr>
<td></td>
<td>Socially responsible restructuring</td>
<td>A sustainable logistics company must massively incorporate AI into many of its services to remain competitive. But it wishes to do so in socially responsible fashion.</td>
</tr>
<tr>
<td></td>
<td>A new committee on professional development</td>
<td>A company’s professional development committee welcomes new members: the representatives of collaborating robots. Not everyone shares the same opinion about this change.</td>
</tr>
</tbody>
</table>
2. Five full scenarios

The five scenarios selected each explore a possible situation in 2025 for one of the themes discussed in the first co-construction phase of the Montréal Declaration: predictive health, predictive education, smart city, predictive justice, and the transversal theme of changes in the workplace.

Each scenario presents the story of a case that was built by integrating numerous dimensions: a sector problem, a user experience set in 2025, a learning apparatus that uses data mobilization and one or more artificial intelligence techniques, and finally, ethical and social issues.

Table 8: Elements of the five scenarios

<table>
<thead>
<tr>
<th>2025 AI Scenarios</th>
<th>Digital twins</th>
<th>Self-driving cars</th>
<th>AlterEgo</th>
<th>Parole</th>
<th>Responsible restructuring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Themes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sectoral issue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preventive and personalized healthcare using similar profiles</td>
<td>Safety and sharing the road</td>
<td>Personalized learning at school</td>
<td>A judge’s decision in the case of uncertainty</td>
<td>Preventive and socially responsible management of changes</td>
<td></td>
</tr>
<tr>
<td>Types of AI learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustering data into homogenous groups through unsupervised learning</td>
<td>Algorithms of self-driving cars for vision, decision-making supervised and reinforced (learning)</td>
<td>Supervised teaching (student concentration) and reinforcement (homework follow-up policy)</td>
<td>Supervised teaching of past cases of recidivism</td>
<td>All AI from the moment it creates changes in companies and administrations</td>
<td></td>
</tr>
<tr>
<td>Ethical and social issues (examples)</td>
<td>Privacy: data confidentiality</td>
<td>Justice: equitable sharing of public spaces</td>
<td>Privacy: confidentiality of student data</td>
<td>Autonomy and critical knowledge in decision-making</td>
<td>Justice: equitable sharing of productivity gains</td>
</tr>
</tbody>
</table>
Theme 1: PREDICTIVE HEALTH

Initial scenario: DIGITAL TWINS

MARCH 10, 2025. Olivier receives a notification on his phone that one of his digital twins has just been diagnosed with depression. Digital twins are people who share the same biological traits and have similar health profiles. All data pertaining to Olivier’s health has been collected by Health Canada since December 2023. Some is provided by his phone’s health app (such as the number of steps taken in a day, or the number of hours of sleep), and from what he shares publicly on social media (data purchased from Alphabet and Baidu). They are cross-referenced with data provided directly from the healthcare system regarding his disease history and genetic predisposition. This data is linked to that of the entire population in the “world health cloud”, overseen by the World Health Organization since 2023, which helps define individual health profiles to offer each person targeted and highly personalized prevention and precision medicine.

Olivier thus discovers that morning that he is at risk of developing the same pathology as one of his 126 digital twins. Faced with this prognosis, Health Canada’s algorithm recommends that Olivier go to a mental health clinic to receive a personalized preventive treatment, reduce his workload to less than 40 hours a week, and increase his physical activity, given the proven beneficial effects of sports to prevent depression. Olivier decides to ignore this advice, as he is working on a contract that could have major repercussions on his career. However, over the course of that week, he learns that 25 of his digital twins have received a similar diagnosis.

Theme 2: SMART CITY

Initial scenario: SELF-DRIVING CAR – SETTING THE ALGORITHM AND SHARING THE ROAD

FALL 2025. The Plateau-Mont-Royal and the Rosemont—La Petite—Patrie boroughs came together to create a pilot project zone in Montréal where priority is given to self-driving electric vehicles.

The self-driving vehicles, privately owned or carsharing (Communauto, Car2go and the new Goober pods) as well as self-driving STM shuttles travel at a speed of 25 km/h to ensure maximum security for users, cyclists and pedestrians (“Zero accident” policy from the City). This policy ensures fluid circulation without traffic jams, with dynamic traffic lights thanks to a network of connected sensors. All this gives users the freedom to take part in activities such as working, writing, or listening to music in their vehicle without being disturbed by jerking movements. Vehicles with drivers must adapt to these speeds, or risk deterrent fines. The new self-driving traffic regulation centre (SDTRC) does, however, authorize a speed of 50 km/h during morning and evening peak hours on certain major roads, such as Papineau Avenue, Iberville Street and Saint-Joseph Boulevard. To ensure the safety of pedestrians and prevent them from crossing these roads in on a whim, safety barriers have also been erected along these roads.

Samia, 30, lives in Rosemont. She’s a massage therapist, strongly suited to therapeutic relationships and an animal rights activist. She lives with her partner, Robin, a computer technician, and her cat, Linus, 4. As often as possible, she lets Linus roam freely throughout town, as she can always track him thanks to his connected collar. The very moderated speed of the self-driving cars reassures her about her cat. Furthermore, she appreciates that in this Montréal pilot project zone, the cars are set in “altruistic” mode, which means they act in the interests of the greatest number of people, even at the expense of the person in the car.
But during the summer, a group of cyclists grows tired of seeing the many safety barriers that appropriate public space for self-driving cars. Since the end of August, they have been protesting by organizing "free bike parades" on the borough’s streets for the sake of sharing the road with all eco-friendly methods of transportation, never hesitating to throw themselves under the wheels of the self-driving vehicles, knowing that their "altruistic setting" saves them from danger. But on this October morning, Samia, in her car, doesn’t know that her husband Robin modified—out of love—the setting in her car to make it "selfish": it now preserves the driver’s interests in the event of an accident. When Laurène, a free bike activist, jumps the security barrier and throws herself in front of the car on Papineau Boulevard, it does not react as planned. An accident occurs that severely injures Laurène, because the CRTA technicians didn’t lower the speed from 50 km/h to 10 km/h when she jumped the safety barrier. Samia is in a state of shock.

This year everything was going to change thanks to AlterEgo, an artificial intelligence that assists teachers. AlterEgo measures in real time the attention span of students, identifies what hinders their understanding during the lesson and detects exceptional students. The device is very simple: thanks to sensors housed in an electronic bracelet that is connected to the tablet on which the student is working, AlterEgo detects the stress felt by children and when they start to lose focus. The device is also able to analyze reading speeds to identify students with comprehension problems.

Today, Carmen gives the students their bracelet and answers questions from parents who have been invited to attend the first class. The parents were initially a little surprised by the device, but they now seem seduced by everything that it can do. The children play with their electronic bracelet and keep asking AlterEgo questions on their tablet: "AlterEgo, who’s your favourite singer?" At the same time, AlterEgo gets acquainted with the students and starts recording the first data.

Carmen explains that her assistant also makes pedagogical recommendations. It can remove parts of the lesson that are deemed ineffective or unsuitable for learning. At the end of the day, Carmen must study AlterEgo’s recommendations and each student’s profile to plan and adapt the lesson. This greatly improves student tracking. "Thanks to AlterEgo, there’s almost no more stress related to exams or evaluating students’ needs and progress!" says Carmen. Student assessment will now be almost continuous. However, Carmen is quick to reassure some doubtful parents: teachers will still be assessing students’ needs and progress. AlterEgo is an addition to that process. “Who will grade the exams? Will AlterEgo do that too?” asks Hourya’s father. Carmen smiles and wraps up her presentation with a joke: “When I have to work at night, I’ll definitely need AlterEgo to take care of my kids, Lola and Emiliano. It just might come true!”

Theme 3: PREDICTIVE EDUCATION

Initial scenario: SELF-DRIVING CAR – SETTING THE ALGORITHM AND SHARING THE ROAD

AUGUST 28, 2025. Carmen starts her third year as a teacher at the Thérèse-Casgrain Elementary School. Like last year, she will be teaching Grade 6. She is eager to use the new teaching methods that the Commission scolaire de la Baie (Baie School Board) has set up a pilot project in the school to improve support for exceptional students and adapt teaching techniques to different learning styles and needs. Last year, Carmen spotted Samuel’s learning disabilities a bit late in the school term. Samuel had attention-deficit problems, chatted with his peers instead of listening and would sometimes act aggressively towards friends. Carmen thought his low grades were related to an attention deficit disorder (ADD). She talked about it with Samuel’s parents. The conversation did not go very well.
Theme 4: JUSTICE AND PREDICTIVE POLICE

Variable scenario: PAROLE DECISION

FALL 2025. Sylvia, 29, has been dating Jean for ten years. When she learned Jean cheated on her, she sought revenge by hacking his connected refrigerator.

Knowing Jean’s severe peanut allergy, his refrigerator, which would send his grocery list to a partner store, would compile the list according to this information. However, once Sylvia hacked the system, Jean’s peanut allergy no longer appeared in the default parameters and the refrigerator produced a list that was no longer adapted to his health requirements. While eating a prepared dish which contained trace amounts of peanuts, Jean started having difficulty breathing and was rushed to the hospital.

Sylvia was arrested for her crime. At the time of sentencing, an algorithm calculated an 80% chance of her relapsing in the next two years, and sentenced her to a two-year prison sentence and a $10,000 fine.

To arrive at this recommendation, the algorithm calculated the risk based on many factors:

> Static historical factors, such as the age at which Sylvia committed her first infraction and her prior offences (Sylvia had already hacked her mother’s pillbox at 18, and her neighbourhood’s video surveillance camera network at 25);

> Dynamic risk factors: Sylvia’s occupation, the company she keeps, her family and romantic relationships, the regret expressed by Sylvia, etc.

Then the algorithm compared Sylvia’s case to many other similar cases.

Following the decision rendered by the algorithm, the judge could choose it or order probation for Sylvia, on the condition she follows an all-new rehabilitation program for delinquents that has no data history, meaning no possible interpretation by the algorithm.

The judge, who is keen on social innovation, chose the second option. The rehabilitation program recommends Sylvia be evaluated and follow a routine personalized plan for two and a half years, as well as find a legal job. Given her hacking skills, Sylvia is also asked to put her knowledge to good use by contributing to the field of cyber security.

Theme 5: WORKPLACE

Initial scenario: SOCIALLY RESPONSIBLE RESTRUCTURING

JANUARY 15, 2025. Created in 2020 in Montréal, Zéro Carbone Logistique (ZCL) is a new world leader in sustainable logistics, and has witnessed incredible growth over the past five years. The company currently employs 3000 people in Montréal.

Ever since it was founded, ZCL has wanted to include its environmental and social objectives in its shareholder agreement by adhering to B Corp status and by following the ISO 26000 standards on the social responsibility of companies. This policy was beneficial for ZCL because many union funds and socially responsible investment funds quickly invested in the company, which became a poster child for green start-ups in Quebec.

However, ZCL is a company that must be profitable, and it faces very fierce competition when it comes to the cost of services: offering environmental value isn’t enough to prosper. Like many companies, it conducted a financial audit and the report strongly recommended a radical scenario to ensure the company’s sustainability: massively investing in AI and the automation of several tasks, starting in 2020. This includes calculating each trip’s carbon footprint, self-driving electric trucks, parcel sorting, routing blimps and electrical boats, as well as administrative follow-up on files. In total, 1000 jobs out of 3000 could be eliminated, and 1000 others must move towards types of cooperation between humans and co-bots! For ZCL management, there’s no way this evolution will be done abruptly, and they wish to establish a “socially responsible restructuring”, by carefully preparing the collaborators for new positions.

---

31 A certification issued to companies that satisfy societal, environmental, governance and public transparency requirements.
Nabila, one of the founders of ZCL, suggests the following solution: creating, in partnership with one of the web giants, a massive data processing platform used by AI applications in logistics. Jean-Raymond, the company’s union representative, is very worried: he mentions that these companies feed off underpaid workers who spend 15 hours a day coding data to train algorithms, and that it is not a good solution for his colleagues. He would rather establish a cooperative data processing platform. “They have some in California and they’re much more in line with our values.” But a big web stakeholder is ready to immediately invest in massive data for sustainable logistics and create, with ZCL, a subsidiary in Montréal that could hire most of the 1,000 people. Time is running out; their investors are pushing for the immediate partnership which is a sure thing, even though it will most certainly have an impact on ZCL’s image. Nabila and Jean-Raymond had been raising these issues with the executive committee on many occasions since 2023. They would have liked to seek advice from a public service earlier, but didn’t know whom to reach out to and now, it’s too late.
The Montréal Responsible AI Declaration was prepared under the direction of:

Marc-Antoine Dilhac, the project’s founder and Chair of the Declaration Development Committee, Scientific Co-Director of the Co-Construction, Full Professor, Department of Philosophy, Université de Montréal, Canada Research Chair on Public Ethics and Political Theory, Chair of the Ethics and Politics Group, Centre de recherche en étique (CRÉ)

Christophe Abrassart, Scientific Co-Director of the Co-Construction, Professor in the School of Design and Co-Director of Lab Ville Prospective in the Faculty of Planning of the Université de Montréal, member of Centre de recherche en étique (CRÉ)

Nathalie Voarino, Scientific Coordinator of the Declaration team, PhD Candidate in Bioethics, Université de Montréal

Coordination
Anne-Marie Savoie, Advisor, Vice-Rectorate of Research, Discovery, Creation and Innovation, Université de Montréal

Content contribution
Camille Vézy, PhD Candidate in Communication Studies, Université de Montréal

Revising and editing
Chantal Berthiaume, Content Manager and Editor
Anne-Marie Savoie, Advisor, Vice-Rectorate of Research, Discovery, Creation and Innovation, Université de Montréal
Joliane Grandmont-Benoit, Project Coordinator, Vice-Rectorate of Student and Academic Affairs, Université de Montréal

Translation
Rachel Anne Normand and François Girard, Linguistic Services
Rebecca Sellers, Copywriter, Translator, ESL Teacher

Graphic design
Stéphanie Hauschild, Art Director

This report would not have been possible without the input of the citizens, professionals and experts who took part in the workshops.