< >

Montréal Declaration
Responsible AI_

</ >

ACCOUNT OF THE DELIBERATIONS

JUNE 2018

MONTREAL DECLARATION FOR A RESPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

ACCOUNT OF THE DELIBERATIONS

JUNE 2018

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF FIGURES AND TABLES

OUR PARTNERS

Université de Montréal

CRE
CENTRE DE RECHERCHE EN ETHIQUE

LAB VILLE PROSPECTIVE

IVADO
INSTITUT DE VALORISATION DES DONNÉES

SAT

MILA

ICRA
Programme
IA et
société

Université de Montréal | design∩société

polEtHics
CHAIRE DE RECHERCHE DU CANADA
ÉTHIQUE PUBLIQUE ET THÉORIE POLITIQUE

CIRANO
Allier savoir et décision
Centre interuniversitaire de recherche en analyse des organisations

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

UNIVERSITÉ
LAVAL
Institut d'éthique appliquée

MUSÉE DE LA
CIVILISATION
Québec

CENTRE D'ÉTUDES
ET DE RECHERCHES
INTERNATIONALES
CÉRIUM
Université de Montréal

crdm_ul
CENTRE DE RECHERCHE
EN DONNÉES MASSIVES
DE L'UNIVERSITÉ LAVAL

CREDITS

This report on the results of the citizen deliberations was written under the direction of

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 (CRÉ)

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 (CRÉ), Canada Research Chair in Public Ethics and Political Theory

Scientific Coordination

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

Assisted by
Camille Vézy, Doctoral student in communications, Université de Montréal

Writing team

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 (CRÉ)

Valentine Crosset, Doctoral student in criminology, Université de Montréal

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 (CRÉ), Canada Research Chair in Public Ethics and Political Theory

Martin Gibert, Ethics Counsellor at IVADO and Researcher at Centre de recherche en éthique (CRÉ)

Vincent Mai, Doctoral student in robotics, Université de Montréal

Christophe Mondin, Research Professional for CIRANO

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

Camille Vézy, Doctoral student in communications, Université de Montréal

Alessia Zarzani, PhD in Planning, Université de Montréal and PhD in Landscape and Environment, Université la Sapienza de Roma

Editing and revision team

Isabelle Bayard, Vice-Rector assistant of Research, Discovery, Creation and Innovation, Université de Montréal

Chantal Berthiaume, Content Manager and Writer

Joliane Grandmont-Benoit, Project Coordinator, Vice-rectorate of Student and Academic Affairs, Université de Montréal

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

Work coordinated by

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

The writing of this report would not have been possible without the contribution of the students and professionals who facilitated the workshops and took notes, nor without the involvement of Marie D. Martel, Professor at the École de bibliothéconomie et sciences de l'information de l'Université de Montréal.

And of course, the writing of this report would not have been possible without the reflection of the citizens, professionals and experts who took part in the workshops.

All of their names can be found at the end of this report.

In this document, the use of the masculine gender was adopted to facilitate reading, and has no discriminatory intention.

SUMMARY

On November 3, 2017, Université de Montréal launched the co-construction process for the *Montreal Declaration for a Responsible Development of Artificial Intelligence (Montreal Declaration).* Eight months later, we present the first results of this citizen deliberation process that is now at the halfway point. It's a very favourable assessment: dozens of events were organized to spark discussions about the social issues raised by artificial intelligence (AI), and fifteen deliberation workshops were held over three months, involving over 500 citizens, experts and stakeholders from all horizons.

The *Montreal Declaration* is a collective work that aims to put AI development at the service of the individual and common good, and guide social change by making recommendations with a strong democratic legitimacy.

The selected citizen co-construction method relies on a preliminary declaration of general ethical principles articulated around fundamental values: well-being, autonomy, justice, privacy, knowledge, democracy and responsibility.

If one of the goals of the co-construction process is to fine-tune the ethical principles suggested in the preliminary version of the *Montreal 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 the human intelligence, to learn from it and replicate it. For example, discovering complex patterns among a large quantity of data, or reasoning in a probabilistic fashion, in order to sort information into categories, predict quantitative data, or aggregate data. These cognitive skills are the basis for other skills such as choosing among several possible actions to reach a goal, interpret an image or a sound, predict a behaviour, anticipate an event, diagnose a pathology, etc. These AI realizations rest on two elements: data and algorithms, series of instructions that perform a complex action.

To concretely discuss the ethical issues of AI, the co-construction method workshop relies on the preliminary version of the *Montreal Declaration*. Schematically, after deciding on the "why?" (which desirable ethical principles should be included in a declaration on the ethics of AI?), it's a matter of prospectively anticipating, along with the participants, how ethical issues around could arise in the coming years, in the fields of health, justice, smart cities, education and culture, the workplace and public services. Then, we imagine how we could respond to these issues. For example, through a measure such as a sectorial certification, a new actor mediator, a form or a standard, through a public policy or research program.

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

Other citizens choose to contribute to the reflection by filling out a questionnaire online or tabling a brief. The results of these specific initiatives will be discussed in the global report on the activities tied to the *Montreal Declaration*, which should be published in the fall of 2018.

Co-construction workshop results – The general trends

Generally speaking, the participants recognized that the arrival of AI came with important potential benefits. Namely, in their field of work, participants recognized the time savings that AI devices could bring. However, it was also mentioned that AI development had to be done with caution and right now to prevent abuse, although some consider the possibilities brought on by AI to still be limited.

The citizens highlighted the need to implement different mechanisms to ensure the quality, intelligibility, transparency and relevance of the information being communicated. They also discussed the difficulty of guaranteeing truly enlightened consent.

The great majority of the participants recognized the necessity to align public interests with private ones and prevent the apparition of monopolies, or limit the influence of corporations through more cohesive and legal measures.

The participants also recommended putting mechanisms in place that would come from and involve independent, trained people to favour the diversity and integration of those who are most vulnerable, and protect the mixed aspect of the lifestyles.

Whatever the use, the majority of the participants insisted on the fact that AI must remain a tool, and that the final decision must come from a human being.

Priorities according to the *Montreal Declaration* principles.

The responsibility principle has often been deemed the most pressing issue, followed by autonomy, privacy, then well-being (individual and collective), knowledge and justice. It's worth noting, however, that they are all closely linked.

As for the autonomy principle, which is often selected as a priority, it has to do with preservation, or even encouraging individual autonomy when faced with the risks of technological determinism and dependency on tools. It also raises the issue of a double liberty of choice: being able to follow your own choice when faced with an AI-guided decision, but also the choice not to use these tools without risking social exclusion.

The well-being principle is also an important one for participants. It is implicit at every table, illustrating a collective wish to move towards a just and equitable society that fosters the development of all.

In a general sense, the well-being principle has also taken on the form of a call to maintain a genuine human and emotional relationship between experts and users in every field.

Issues that can lead to the creation of new principles, or new themes to explore and deliberate.

The impact of the responsible use and development of AI on the **environment** raises issues, namely on the way to guarantee the responsible and equitable use of material and natural resources.

The justice principle was discussed on the basis of two types of issues, which could lead to two new principles: a **diversity principle** looking to avoid discrimination by finding bias-free mechanisms and an **equity or social justice principle**, which would require AI benefits to be accessible to all, and that the development of AI not contribute to the growing economic and social inequalities, but rather help bridge the gap.

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

A transparency principle. This principle implies being able to understand an algorithmic decision and react to it. For this, citizens think it's important that the algorithmic procedures be explainable so they can see and understand which criteria were considered in the decision.

Whatever the field, the citizens identified many issues regarding the relationship between human beings and AI.

Whatever the field, the citizens identified many issues regarding the relationship between human beings and AI.

Potential solutions

All the co-construction tables agreed on **3 potential solution**s to guarantee socially responsible AI development, regardless of the field:

1. Legal provisions;
2. Training offered to all and
3. The identification of key independent players for AI management.

Continue the Deliberation

The *Montreal Declaration* project concentrated its first phase on five key sectors: education, health, work, smart city and predictive police. An entire year of co-construction wouldn't even cover all the reflection themes. The co-construction initiative will therefore continue in September 2018, allowing for discussions about new themes that had barely been touched upon in the scenarios used in the co-construction phase. For example: environment, democracy and media propaganda, as well as security and integrity.

We will present public policy recommendations around priority fields of action. To date, we can say that three fields of action have established themselves: digital literacy, diversity and inclusion, and transition and social mutations. The final results will be presented in December 2018.

INTRODUCTION

On November 3, 2017, Université de Montréal launched the co-construction process around the *Montreal Declaration for a Responsible Development of Artificial Intelligence (Montreal Declaration).*

We had no idea of the interest the initiative would capture, nor of the size of the task that lay ahead. Eight months later, we present the first results of the citizen deliberations, halfway through the process. It is a very successful one: dozens of events were held to discuss the social issue surrounding AI, and fifteen deliberation workshops took place over three months, involving over 500 citizens, experts and stakeholders of all professional horizons.

The halfway report we are presenting must be taken as a temporary, non-exhaustive summary of a democratic deliberation process to enlighten public policy decisions regarding artificial intelligence. The work around what we call the *Montreal Declaration* was led by a multidisciplinary and inter-university team of researchers, mainly in Quebec but also across the world. Awareness of the social issues around artificial intelligence is shared by this research community, but also by society as a whole. We therefore suggested a citizen co-construction process because we believe 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 carries out a prospective design of applied ethics, because it's a matter of anticipating ethical controversies around future artificial intelligence technologies or social situations where the use of these technologies is pushed to the limit of what we can anticipate. Then, we carried out this consultation process on an unheard-of scale. The numbers mentioned above paint a clear picture. This process, to be clear, will continue, and as the *Montreal Declaration* remains open to review, the co-construction will not end when this first deliberative endeavour does.

We called the public around the drafting of the Declaration, and were called upon in return: what will the Declaration change? Who is writing? Isn't this just a vain university professor thing? Aren't there already too many manifestos, professions of faith on the ethical values of artificial intelligence? Isn't surrounding the development of artificial intelligence with ethical principles and recommendations a way of condoning it? Isn't that approving a technocratic vision of society? Why not devote our energy to criticizing this development? None of these interrogations are without merit, and because we are committed to increased transparency around artificial intelligence, we are also committed to increased transparency around human and collective intelligence. This halfway report will, we hope, provide a few answers.

The ethics of artificial intelligence have been a hot topic in many countries over the last two years. Every actor in its development, researchers, businesses, citizens, political representatives, all recognize the urgency of establishing an ethical, political and legal framework to guide the research and use of artificial intelligence. Because there is no doubt that we are at the dawn of a new industrial revolution with the rise of artificial intelligence technologies. The impacts of this revolution on the production of goods, the delivery of services, the organization of work and the workforce, or even on family and personal relationships are still unknown but will be major, possibly disruptive in certain fields. Indeed, the societal changes brought about by artificial intelligence are surprising in their suddenness and spark varied reactions, from enthusiasm to disapproval and scepticism. We could ignore them and launch into speculations around the existence or not of that we call artificial intelligence, but we'd only be pushing back the problem to a time when it will no longer be possible to influence its development.

Many objections and fears were raised during this first co-construction process. Many workshop participants and observers in the Declaration project questioned the technocratic ideology that sees in technology a way of rationally organizing all of society, and that reduces social issues to technical problems. Others question the ability and the will of public institutions to regulate lucrative technologies. These objections must not be casually dismissed, because they are based on historical precedents that shook their faith in technological innovations, and even more so in the people promoting them. But it is also important that those raising objections don't undermine every effort to positively influence the future of society and support them by getting involved in the democratic deliberations that allow us to keep control. We can complain about the effects of new information technologies and artificial intelligence on social relationships, we can criticize the reduction of social life to a series of lifestyles, this will not prevent technological innovation, nor will it influence it. Yet that is the entire purpose of the *Montreal Declaration*: guide the development of artificial intelligence in order to promote or preserve fundamental ethical and societal interests.

In conclusion, we will not settle the unrelenting question regarding the use of the term "artificial intelligence": is it appropriate to refer to data processing, recognition and decision-making algorithms? Its use can be contested by opposing the fact that artificial intelligence refers to very limited knowledge processes when compared with human intelligence, or even the intelligence of pigeons. It's undeniable. But with that reasoning, paramecia offer complexity that surpasses that of any algorithm, even a learning one. By going down that path, you come across a deadlock of intelligence as a whole. What is human intelligence? The hundreds of thousands of pages that have been written to answer that question still doesn't suffice.

However, a few statements can help avoid misunderstandings that are at the root of the controversy: firstly, people often confuse intelligence and thought. Intelligence is a property of thought, it is not thought as a whole. Then, intelligence is particular in that it reduces the complexity of the world in which the intelligent being evolves to allow him to better master his environment. We give ourselves rules to analyze reality, calculate it, evaluate it and make decisions. A long philosophical tradition of thinkers that did not lack intelligence have claimed it from Socrates to Rusell, along with Leibniz. I a certain way, intelligence reduces reality to better act on it. Finally, stemming from the above, intelligence, even human, is largely algorithmic: it analyzes data and makes calculations according to procedures. Sometimes these procedures are inadequate, the analysis is wrong. But to establish it in the first place, you must first analyze and use procedures.

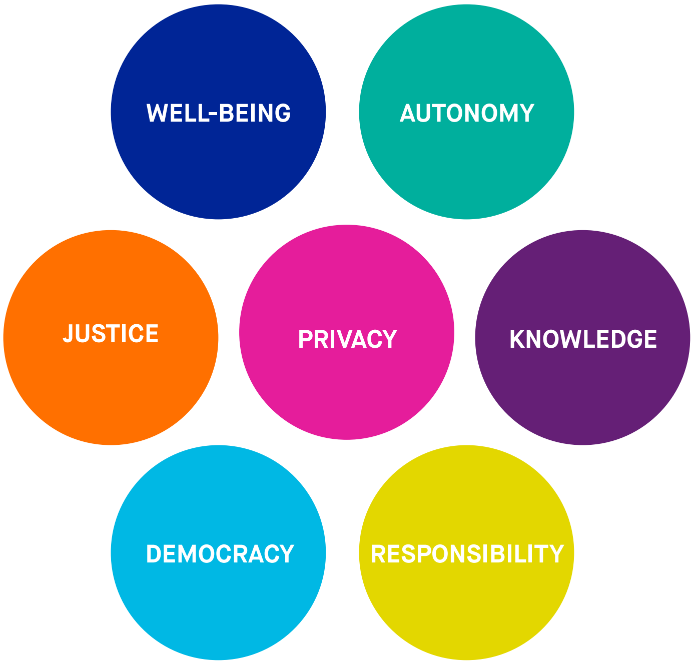
Reflecting on the goals we wish to pursue is not strictly a matter of calculations. Directing your personal and social life towards certain worthwhile goals does not depend on an algorithmic procedure. Knowing if we must use nuclear weapons to kill the greatest number of people and weaken an enemy country cannot be solely determined by a calculation of the consequences. There's something tragic about avoiding reflection on moral consequences by seeking only a calculation of the means. That being said, it's true that artificial intelligence cannot do it, and if it could do it we'd have another set of problems facing humanity's future, much more pressing than those we are faced with today. In the world we know and can anticipate in the near and mid-term future, the reflection on the finality of social life and existence in general is still a product of human intelligence.

The *Montreal Declaration* rests entirely on this statement: it is up to human and collective intelligence to define the purposes of social life and, accordingly, the direction of artificial intelligence development so that it is socially responsible and acceptable, even desirable.

1. WHY HAVE A MONTREAL DECLARATION RESPONSIBLE AI?

The *Montreal Declaration* is a collective work that aims to put the development of artificial intelligence to work for the good of everyone, and orient social change by developing recommendations with a strong democratic legitimacy.

The selected method of citizen co-construction rests on a preliminary declaration of general ethical principles that state **FUNDAMENTAL VALUES.**



The initial work of identifying these values and principles allows us to launch a citizen involvement process that will define the ethical principles of responsible AI development and the recommendations to put into place to ensure that AI is promoting fundamental human interests.

1.1

THE INTELLECTUAL ORIGINS OF THIS PROJECT

The artificial intelligence (AI), and more specifically deep learning, revolution opens perspectives to unimagined technological developments that will help improve decision-making, reduce certain risks and offer assistance to those who are most vulnerable. This revolution is remarkable in many ways, although it also revives questions that were first raised in the 18th century, at the time of the Industrial Revolution. It would be unwise to ignore the unique aspect of this revolution by hiding behind platitudes that aren't preparing us 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 the trend to automation—but upon closer inspection nothing is similar to what's in play with the arrival of artificial intelligence technologies. The cognitive skills we believed unique to humans can now be exercised by algorithms, machines that must be recognized as, in a certain sense, intelligent.

The social impacts of these new technologies, although very diverse, are still somewhat unknown. They could prove brutal if we don't take the time now to have an ethical, political, judicial, sociological or psychological reflection on the type of society and human relationships we want to promote or protect while still using the advantages of the information technologies and algorithm calculations.

The use of algorithms to make technical or administrative decisions isn't new. The rise of decision-making algorithms truly begins in the 1950s, especially in the healthcare field: emergency room triage in hospitals, detection of sudden infant death syndrome risks, heart attack prediction[1]. All these algorithm techniques, "the procedures" already raise a certain number of ethical and social issues: those of social acceptability of an "automatic" decision, of the final decision (is a human being at the end of the decision-making

[1] Paul Meehl, *Clinical versus Statistical Prediction*, University of Minnesota, 1954.

chain?), or of responsibility in case of a mistake. And it is clear that these issues are being raised again with the latest algorithmic innovations.

What is different, then, about the latest technologies that full under the AI acronym? From an objective standpoint, what changes is the quantity of information that can be handled by computers (massive data) and the complexity of learning algorithms that, by feeding off of massive data, can accomplish perceptive and cognitive tasks allowing visual or audio recognition, and make decisions in defined contexts. By combining different features (facial recognition, behaviour analysis, decision-making), AI raises extremely important ethical problems. From a subjective point of view, what's new is the citizen wake-up call, as late as it was sudden, around the issues of algorithmic governance, the treatment of personal data and the social impact that some professional sectors are already experiencing.

If the progress of AI can surprise and fascinate, it also awakens the fear that using machines, namely robots, will considerably reduce the human relationship component when it comes to medical treatment, elderly care, legal representation, or even teaching. The reactions to the development of artificial intelligence can even prove to be hostile when AI is used for increased control of individuals and society, a loss of independence and a curtailing of civil liberties. This is why the hope of artificial intelligence being the bringer of social progress, always holds a dark shadow: placed into 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 the researchers, the developers, the entrepreneurs and the policymakers.

The development of AI and its applications therefore involves fundamental ethical values than can come into conflict and create serious moral dilemmas and deep social and political controversies: must we prefer public safety by increasing smart surveillance (facial recognition, anticipating violent behaviour) at the expense of individual freedoms? Objectively improving the well-being of individuals, namely by encouraging people to adopt behaviours normalized by smart devices (nutritional behaviour, work management, day planner) can it be done while still respecting people's independence? Should the economic performance targets take priority over a concern for an equitable share of the benefits of the AI market?

These dilemmas or tensions cannot be overcome simply by ranking fundamental values and interests. To put it another way, it's not about classing the values in order of importance a priori, or building a simple and unequivocal scale of values, let alone favouring some while ignoring others (security at the expense of liberty, efficiency without social justice, well-being at the expense of independence). We also can't hope to find unique and permanent solutions. It's better to take the moral dilemmas caused by the development of AI seriously and collectively build an ethical, political and legal framework that will allow us to fact it while respecting the different fundamental values that we legitimately hold as members of a democratic society.

1. 2

FORUM ON THE SOCIALLY REPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

These reflections were the starting point for the initiative by the Fonds de recherche du Québec et de l'Université de Montréal to organize an international meeting to discuss the social impacts of AI and start the work around the *Montreal Declaration for a Responsible Development of Artificial Intelligence*[2]. On November 2 and 3, 2017, at the Palais des congrès de Montréal, a forum bringing together the greatest experts in the fields concerned by a reflection on AI, from pure science to social sciences and humanities. The Forum suggested setting the guidelines for a collective reflection on the ethical land socially responsible development of artificial intelligence, by pursuing the following three objectives:

> offer a public reflection space around AI development issues and its social impacts;

> raise interest and notoriety among decision makers, industrial partners, politicians and the general community interested in AI, making them aware of the social questions raised by the sudden growth and numerous uses of AI;

> privilege an interdisciplinary and intersectorial approach as a key factor to successful ethical and sustainable AI.

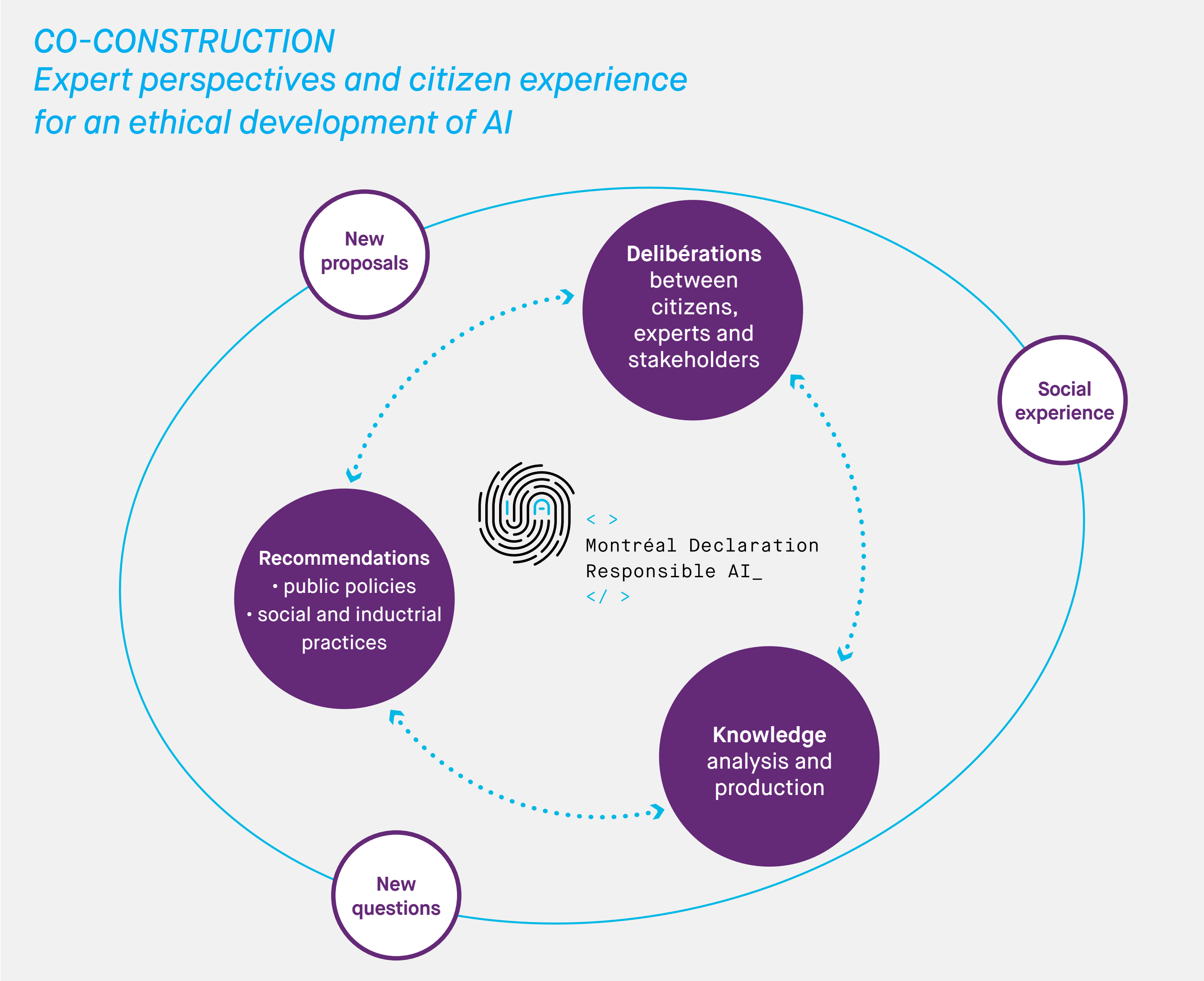
Thus were defined the guidelines on an inclusive approach (interdisciplinary and intersectorial) which is at the heart of the elaboration of the *Montreal Declaration for a Responsible Development of Artificial Intelligence* that is not only responsible, but socially progressive, guaranteeing equality and justice. The preliminary version of the *Montreal Declaration* was presented at the end of the Forum. It was then a matter of launching the citizen co-construction process around AI ethics, a process we will expand upon in section 4.

[2] 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, Vice-rectorat à la recherche, à la découverte, à la création et à l'innovation), Gregor Murray (Université de Montréal, École de relations industrielles et CRIMT – Centre de recherche interuniversitaire sur la mondialisation et le travail), Doina Precup (Université McGill, School of Computer Science; MILA), 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).

1. 3

TOWARDS THE *MONTREAL DECLARATION*

Figure 2 : The co-construction approach


CO-CONSTRUCTION
Expert perspectives and citizen experience
for an ethical development of AI
New proposals
Delibérations between citizens, experts and stakeholders
Social experience
Recommendations
• public policies
• social and inductrial practices
Montréal Declaration
Responsible AI_
Knowledge analysis and production
New questions

The initial identification of these values and corresponding principles was only designed to launch the citizen participation process that will help fine-tune the ethical principles of responsible AI development, add to them and complete them. It should therefore come as no surprise that the *Montreal Declaration* is schematic and that the statement of principles is willingly very simple and consensual, leaving the interpretation and completion to the public deliberations[3].

[3] 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 is to fine-tune the ethical principles suggested in the preliminary version of the *Montreal Declaration*, another goal, just as important, was developing recommendations to oversee AI research and its industrial and technological development. However, it is too frequent to see analysis reports and recommendations forgotten as soon as they're published: this is why it's crucial to keep up the momentum built in the co-construction period.

Once the co-construction process is complete, it is necessary to open a public debate in the arenas where political, legal and policy decisions are made, in order to concretely implement the recommendations that came out of the citizen deliberation. These recommendations are not only legal in nature and, when they are, they don't necessarily involve modifying a law. They could, however, request a modification to the legal framework and, in certain fields, they have to. In other cases, the purpose of the recommendations is to nourish and guide the reflection of professional organizations so they modify their code of ethics or adopt a new ethical framework.

This step is therefore the ultimate goal of the co-construction process. We must, however, immediately inform you that when faced with a technology that has never ceased to evolve over the last 70 years and whose major innovations now come every 2 to 5 years, on average, it would be unreasonable to present the *Montreal Declaration* as definitive and complete. It is essential to think of co-construction as an open process, with successive and cyclical deliberation, production and recommendation production stages, and think of the Declaration itself as a guiding document that can be reviewed and adapted according to the evolution of artificial intelligence knowledge and techniques. This process of knowledge production, citizen deliberation and ethical framework and public policy recommendations, will have to be extended into a perennial institutional structure that allows it to remain reactive to AI evolution.

1.4

MONTREAL AND THE INTERNATIONAL CONTEXT

The *Montreal Declaration* initiative is part of a favourable scientific, social and industrial context. Montreal has become a major artificial intelligence research and development hub, with a community of researchers (Yoshua Bengio at Université de Montréal, a pioneer in the field of deep learning, Joëlle Pineau at McGill, and so many others), world-renowned university labs (MILA, IVADO) and an incubator full of thriving start-ups and businesses (Element AI, Imagia to name just a few). This scientific, technological and industrial development is at the heart of a revolution transforming social practices, business models and lifestyles, affecting all sectors of society. The City of Montreal is also this living lab of social and technological change. With fundamental scientific research come social and ethical responsibilities that the Montreal AI community fully accepts.

But outside of Montreal, it's all of Quebec, and all of Canada, that offers a favourable social context to engage in a reflection on the social impacts of AI. Like MILA in Montreal, Vector in Toronto, AMII (Alberta Machine Intelligence Institute) in Edmonton, and the CRDM (Centre de recherche en données massives) in Québec make up hubs of excellence in fundamental research that have brought about extremely quick and robust industrial growth. The Canadian Institute for Advanced Research (CIFAR, or ICRA) played a lead role in this Canadian development of AI by supporting fundamental research when AI was going through its "winter". The *Montreal Declaration* initiative is supported by various players in Quebec and Canada outside of Montreal: the Fonds de recherche du Québec, the CRDM de l'Université Laval à Québec, the Canadian Institute For Advanced Research.

Many international commentators have also shown their support for the *Montreal Declaration*, namely for its elaboration method. The Declaration team was able to establish a dialogue with institutions such as the Royal Society du Royaume-Uni[4] and the EGE (*European Group on Ethics in Science and New Technologies*[5]) that had their own study program and recommendations on AI. We first note a convergence in the guidelines for ethical AI development as well as a shared intent to promote a democratic conception of AI use for the common good.

The *Montreal Declaration* initiative must also be viewed through the international context of an **AI spring**. It is preceded by many initiatives that must be recognized because they catalyzed the reflection around responsible AI. We must first recall the creation, in 2014, of the Future of Life Institute that produced the Asilomar Declaration in 2017: after a 3-day conference, a declaration containing 23 fundamental principles surrounding AI research and its uses was signed by more than 1200 researchers. Professor Yoshua Bengio took part in the event at the time and brought attention to the risks of irresponsible and malicious AI use[6].

Since the Asilomar Conference, many reports on AI ethics have been published. The report from the Association internationale des ingénieurs électriciens et électroniciens (IEEE), *Ethically aligned design*. V2, was made public at the end of 2017 and gathered 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 impacts of AI[7]. Two ambitious strategic reports were published in March and April of 2018: the Mission Villani report in France and the one from the United Kingdom House of Lords "AI in the UK: ready, willing, and able?" Without claiming completeness, let us at least mention 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 an evocative title: "Comment permettre à l'homme de garder la main? – Les enjeux éthiques des algorithmes et de l'intelligence artificielle", in December 2017.

How does the *Montreal Declaration* position itself in this concert of independent initiatives? And what to think about the ethical inflation around AI? This last question is all the more important that we share the same warning as the EGE in its report *Artificial Intelligence, Robotics and "Autonomous" Systems* (March 2018) that in the absence of a coordinated reflection on the ethical and social issues of AI, there exists a risk of "ethics shopping"[8]. The immediate consequence is a form of delocalization of ethical costs in areas of the world where ethical criteria are low priorities. Another risk is a form of trivialization of ethical discourse.

The specificity of the *Montreal Declaration* initiative is that it is essentially participative. From February to April 2018, the co-construction process brought together over 500 citizens, experts and stakeholders over fifteen workshops and co-construction days. Although other participative initiatives have been led elsewhere, namely in France, the *Montreal Declaration* stands out by its size and its prospective methods.

The Montreal Declaration's vocation is to open a dialogue space in Quebec and Canada and offers a collective thinking platform that extends beyond the Canadian borders. The goal is to identify socially acceptable and innovative AI trends using informed citizen reflection in the different concerned democracies as a reference point. This dialogue space must also be accessible to citizens in non-democratic societies that wish to take part in a global debate on the future of human societies.

[4] We wish to thank *UK Science and Innovation Network in Canada* who facilitated the dialogue.

[5] The European Group on Ethics in Science and New Technologies (EGE) is an independent advisory body of the President of the European Commissions.

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

[7] AI Now Institute, "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability", April 2018.

[8] EGE, *Artificial Intelligence, Robotics and 'Autonomous' Systems* (March 2018), p. 14.

2. THE PRELIMINARY VERSION OF THE MONTREAL DECLARATION FOR A RESPONSIBLE DEVELOPMENT OF AI

< >

Montréal Declaration Responsible AI_

</ >

THE MONTREAL DECLARATION FOR A RESPONSIBLE DEVELOPMENT OF AI

(PRELIMINARY VERSION)

PREAMBLE

Intelligence, whether it be natural or artificial, has no value in and of itself. An individual's intelligence does not tell us anything about his or her morality; this is also the case for any other intelligent entity. Intelligence can, however, have an instrumental value: it is a tool that can lead us away from or towards a goal we wish to attain. Thus, artificial intelligence can create new risks and exacerbate social and economic inequalities. But it can also contribute to well-being, freedom and justice.

From an ethical point of view, the development of AI poses previously unknown challenges. For the first time in history, we have the opportunity to create non-human, autonomous and intelligent agents that do not need their creators to accomplish tasks that were previously reserved for the human mind. These intelligent machines do not merely calculate better than human beings, they also look for, process and disseminate information. They interact with sentient beings, human or non-human. Soon, they will even be able to keep them company, as would a parent or a friend.

These artificial agents will come to directly influence our lives. In the long term, we could create "moral machines", machines able to make decisions according to ethical principles. We must ask ourselves if these developments are responsible and desired. And we can hope that AI will make our societies better, in the best interest of, and with respect for, everyone.

VALUES AND PRINCIPLES

WELL-BEING

The development of AI should ultimately promote the well-being of all sentient creatures

AUTONOMY

The development of AI should promote the autonomy of all human beings and control, in a responsible way, the autonomy of computer systems.

JUSTICE

The development of AI should promote justice and seek to eliminate all types of discrimination, notably those linked to gender, age, mental/physical abilities, sexual orientation, ethnic/social origins and religious beliefs.

PRIVACY

The development of AI should offer guarantees respecting personal privacy and allowing people who use it to access their personal data as well as the kinds of information that any algorithm might use.

KNOWLEDGE

The development of AI should promote critical thinking and protect us from propaganda and manipulation.

DEMOCRACY

The development of AI should promote informed participation in public life, cooperation and democratic debate.

RESPONSIBILITY

The various players in the development of AI should assume their responsibility by working against the risks arising from their technological innovations.

3.
THE ETHICAL AND SOCIETAL ISSUES OF AI

The collective reflection process at the heart of the development of the *Montreal Declaration* rests on a preliminary version of the Ethical principles Declaration itself and informative exposes on AI and the ethics of AI.

3.1

WHAT IS AI?

The idea of AI is not a new one. You have to go back to at least the 17th century and the idea of a universal characteristic and combinatorial art from philosopher and mathematician Leibniz: reasoning comes down to calculating, and thought is conceived in algorithmic fashion[9]. The notion of *calculus ratiocination* (logical calculation) predates the idea of an intelligent machine as it will be developed three centuries later, in the 1940s, by Alan Turing. In 1948, in a report entitled "*Intelligent Machinery*" and in 1950, in his famous article "*Computing Machinery and Intelligence*"[10], Alan Turing brings up 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 months, 10-man study of artificial intelligence". But the uses and the development possibilities seemed very limited then, and so began the winter of AI, 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 merely quote hit movies), research in the field never ceased, and the dawn of the 21st century brought springtime for AI.

AI consists of a certain way of simulating human intelligence[11], taking inspiration from it and reproducing it. But foremost it is the brain, the human intelligence headquarters, which was designed as a machine capable of gathering, spotting and collecting data from its environment that it will then analyze, interpret and understand, using this experience to establish connections. The field of AI research consists of producing mathematical tools to formalize the operations of the mind and thereby create machines that can accomplish more or less general cognitive tasks, associated with natural human intelligence. For example, discovering complex patterns among a large quantity of data, or reason in probabilistic fashion, in order to classify according to information 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, interpret an image or sound, predict behaviour, anticipate an event, diagnose a condition, etc.

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

These AI accomplishments rest on two elements: data and algorithms, meaning 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 how to use them correctly (the algorithm). Up until now, data processing capacities (quantity of data and processing algorithms) were too limited to consider a useful development for AI technologies. Things changed with the use of materials that made building very small and very fast

[9] Leibniz (1666), *De Arte combinatoria*.

[10] A. M. Turing (1950), « Computing Machinery and Intelligence ». *Mind* 49, p. 433-460.

[11] 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. »

calculators (computer chips) and storing massive amounts of data possible, as well as with the dawn of the information era thanks to 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's a multitude of inanimate objects, spaces or receivers that are constantly producing unstructured data, meaning coming from disorganized information supports, which must be manipulated and transformed before it can be used. 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 a huge number of images.

But what changed is also 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. Indeed, 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, who also beats any other machine that doesn't use deep learning, and the game of Go, which was reputed unmasterable for algorithms, but which saw AlphaGo triumph over the pest players in the world in 2015.

Although these examples are telling, AI can also serve other purposes such as automating tasks that required human intervention, especially perception and recognition duties. For example: processing speech; recognizing objects, words, shapes, and text; interpreting scenes, colours, similarities or differences in large sets, and by extension data analysis 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're used to wearing small connected objects (phones, watches) and we're 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, etc.

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[12]. 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, my name is Inès. How can I help you shop today?"

"Hi Inès"

For a few moments, you get the impression that there's someone, named Inès, is behind the screen talking to you; for a few moments, it's okay to doubt. 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 with you in a natural fashion. Inès is a conversational agent, a chatbot, AI. It's become commonplace to chat online with chatbots to get more information about your health plan or new bank account, or even ask for 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, it could be considered that the machine successfully passed the Turing test, and we would then be faced with, according to this test, 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?"[13] 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 asking 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 that the interrogator can't tell whether the human being or the machine replied, we can consider that the machine thinks. This is what is meant by "Turing test".

This imitation game caused a lot of controversy and saw philosophers fiercely clash with one another over whether a machine could be said to think. An experiment known as the "the Chinese chamber"

[12] Stanley Kubrick masterfully captured (and helped craft) this fantasy with the very human computer HAL 9000, in his film 2001: *A Space Odyssey* (1968).

[13] A. M. Turing (1950).

was made popular in the 1980s by philosopher John Searle[14]. According to Searle, a machine that outwardly acts in the same fashion as a human being cannot be considered to have intelligence in the strong 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 is far from being settled, but we don't really need to answer Turing's question to wonder about the place AI holds in our lives and in our societies. For now well-trained chatbots can be as good 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 therefore 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 brought about by the introduction of AI into our lives are of a practical nature, whether ethical, political or legal.
It is foremost a questioning of the values and ethical principles, public policy orientations and applying standards surrounding AI research and its uses.

But 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 determining which uses and goals are ethical, socially responsible, and compatible with democratic values and political principles. However, this ethical reflection doesn't 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 did have that goal, however. We must therefore pay close attention to the direction AI research takes, both in universities and developed by private corporations or government organizations.

[14] J. Searle (1980), 'Minds, Brains and Programs'. *Behavioral and Brain Sciences 3*, p. 417–57.

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 experience the limit of their field of intervention. To answer it, we must first very briefly present the field of ethics when discussing the societal issues of AI.

To keep it simple, 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 all 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 demoralizing 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 deception? What must be done in that 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 object isn't the same, nor is the reflection context. Public ethics is concerned with all the 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 disease's progress? 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 various forms of consultations, deliberations or democratic participation, and that is open to a diversity of individual and institutional players such as professionals in the field, association or union representatives, experts, policymakers and citizens. Public ethics call for a collective reflection to establish the principles of the best practices and demands that the players justify their suggestions on the basis of acceptable arguments in a pluralism context. In the case of the medical lie, you can appeal to shared values such as independence, respect of people, dignity, the patient's health or well-being, etc. From these values, it is possible to establish principles that guide the practice of medicine and provide paths to regulation through the implementation of a code of ethics, by modifying a law or enacting a new law.

Public ethics is not besides or above the law, which has its own logic, but it helps clarify the issues of social life that various actors must keep in mind to respond to a citizen's 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, etc.

In the field of AI, it's this type of ethical reflection that we implement. Let's take the example of Melody, a medical conversation agent. Melody makes online diagnoses, accessible on your cellphone, according to the symptoms you describe. In a certain way, it acts as 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 weren't readily apparent with Inès, the shopping advisor chatbot. For example, we should question if Melody must give its user every possible prognosis, even if he is not equipped to understand the information. This problem is a simple transposition of a medical ethic 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 helped clarify a doctor's obligations. Does this solve the problem that Melody and its sister applications that often multiply unchecked?[15] Generally speaking, probably, but specific attention paid to this technology reveals that it's 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.

Other issues are specific to AI and have yet to find ethical solutions. For example, if Melody makes a wrong diagnosis and the condition of the user who followed her advice goes seriously downhill, who is responsible? In the case of a medical consultation with a human doctor, it's very easy to determine who is responsible for a medical error, but that's 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, isn't it rather the certifying body that should be blamed and legally sanctioned?

Public ethics questioning clearly introduces a reflection on the institutions that allow credible responses to be offered 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, we cannot elude the question of the use of developing such intelligent machines, from a social and human standpoint. We must indeed question whether it is acceptable that smart apps replace medical doctors, even accepting the hypothesis that they can make a precise diagnosis, even more precise 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? It is not a "utilitarian" type of question, but a question on the significance of our social relationships, on the recognition of our vulnerability as patients, on human identity. Let's go 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 indeed consider that we should improve access to an efficient public healthcare system and therefore further invest in the training of doctors and an equitable health organization.

[15] 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.

3.4

AI ETHICS AND THE MONTREAL DECLARATION

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

The *Montreal Declaration* is currently in a preliminary version that serves as a starting point to the ethical reflection. The values presented in this version, although incomplete, supply a basic moral vocabulary to begin the ethical analysis of everyday situations and facilitate deliberations. The analysis of the Melody chatbot case illustrates this purpose of the Declaration. To understand the issue of enlightened patient understanding of a diagnosis, of attributing fault in the case of an erroneous diagnosis or of accessing health services, the *Montreal Declaration* offers a list of values you can immediately refer to: autonomy, responsibility, justice. It would be easy to demonstrate that the privacy value helps frame the problem of patient data confidentiality.

The previous sections were presented in the various co-construction days and workshops[16]. They served as a starting point for the deliberations on prospective scenarios and the elaboration of ethical, political and legal solutions to AI societal issues. These informative presentations ended with the following general question:

How must we organize society to make an ethical use of intelligent machines, compatible with our fundamental moral and social interests? Which rules must be followed to make the best use of these machines while protecting our autonomy, ensuring social equality and equitable distribution of the fruits of the AI economy?

The co-construction process helped bring different credible responses to this question.

[16] Other analyses were presented, namely in the case of self-driving vehicles. We will expose them in more in-depth fashion in the final version of this report.

4.
THE
CO-CONSTRUCTION
APPROACH

4.1

THE CO-CONSTRUCTION APPROACH PRINCIPLES

To answer the many interrogations raised by the use of intelligent machines and ensure that AI develops "in good intelligence" with democracy, it is necessary to use an "excess" of democracy and involve the greatest number of citizens in the reflection process around 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 responsibly use AI.

It's not only a question of knowing what people think of a certain innovation and surveying their "intuitive" preferences; co-construction is not a public opinion survey around questions such as: "Are you scared that AI will replace judges?", "Would you prefer that a human or a robot operate on you?" This type of question is not without interest, and the survey method provides important information do policymakers, as well as important working material for social sciences. However, although co-construction invites collective reflection around democratic issues, it also requires the development of documented, credible answers to pressing questions and the formulation of political and legal recommendations benefitting from a strong democratic legitimacy.

This is the entire reasoning behind the approach initiated by the *Montreal Declaration*: giving back to democracy the ability to settle moral and political questions that concern 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

From 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 disqualify public involvement:

1. Ignorance: according to this objection, which is the most common, the public is ignorant and does not possess the ability to understand complex issues that require scientific knowledge, mastery of logical forms of argument and knowledge of political and legal processes.

2. Populism: according to this objection, which is a frightening one, the involvement of the unqualified public can be an opportunity for the demagogic manipulation that stokes popular stereotypes and can lead to the passing of unreasonable propositions, hostile to social progress, or even 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 leads them towards 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't thought of, or were unable to support publicly.

If, for certain individuals, prejudices and a tendency towards irrationality cannot be completely eliminated, it is possible to overcome these biases collectively. In favourable conditions, non-expert individuals can take part in complex debates surrounding social problems, such as those presented today by artificial intelligence research and its industrial applications. Experts in various matters relevant to citizen involvement on artificial intelligence can help implement these favourable conditions.

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

Epistemic diversity

We must first ensure that the deliberating groups are as diverse as possible, in terms of social environment, sex, generation or ethnic origin. This diversity is not only required by the idea we have of an inclusive democracy, but is also necessary 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 discussion.

Access to relevant information

But we know, however, that epistemic diversity is not enough and that if the participants have no skills or knowledge in the field being discussed, they cannot produce new knowledge, or find their way in the discussion. They are then collectively susceptible to amplify individual mistakes. We must therefore prepare the participants by providing relevant, quality information, both accessible and reliable. The deliberations must therefore be preceded by an information phase.

Moderation

Other than having quality information, it is necessary that the participants reason freely, which is to say 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 most problematic in a deliberation is the confirmation bias: we have a tendance to only accept opinions that confirm our own beliefs, and to reject those that go against what we already believe. There are dozens of cognitive biases that can deform the logic course of our reflections.

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 artificial intelligence innovations, it is quite probable that they will be entirely hostile towards them at the end of the deliberation process. To avoid this type of knee-jerk reaction, we believe it's important to ensure epistemic diversity in the deliberating group and to put a moderation body in place.

This does not necessarily have to take the shape of a personal intervention from a moderator. Although we don't reject personal moderation, we believe we can overcome deliberation biases through other means, such as introducing unexpected events in the scenarios that sparked the discussions.

Iteration

Ideally, we should be able to bring together the population as a whole to take part in a reflection on the responsible development of artificial intelligence. But the conditions we just described cannot be applied to very large groups, let alone a society of many millions of people. It is therefore important to conduct citizen involvement in smaller groups and increase the number of meetings. This is the iteration phase of co-construction.

The reasons to proceed this way are technical, but easily understood. The mathematician and player in the French Revolution, the marquis de Condorcet, had demonstrated that the judgment of groups is always right more often than each person individually, and that this increases as the group grows larger. For this to be the case, however, two conditions must be met: the individuals in the group must have more than a fifty/fifty (50/50) chance of being right, and they must not communicate with one another (Condorcet rightly feared the risks of manipulation).

Yet we cannot ensure that for very large groups that the individuals have the required skills and that each individual has more than a fifty-fifty chance of having an appropriate opinion. Allowing deliberation (communication between one another) is one way to increase the skill of the participants, as long as it is in the framework we are suggesting. Of course, that does not satisfy Condorcet's second condition, but it does guarantee the first. And to increase the quality of opinions, it is necessary to multiply the deliberating groups: since we cannot increase the size of the group, we must increase the number of participants by proceeding with an iteration of participation sessions.

For all of these reasons, we have selected the structure of a co-construction workshop that brings together non-expert citizens, experts, stakeholders (associations, unions, professional representatives, businesses), as well as political players. These workshops are organized in different formats adapted to the deliberation spaces and satisfy the conditions for fruitful, solid citizen involvement.

4.1.2 Experts and citizens

Why allow citizens to be heard on complex ethical and political questions that require a good knowledge 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 orientations of its development.

Even when we are not in the presence of a dilemma, strictly speaking, public ethics questions cannot be solved without making choices that favour certain moral interests over other moral interests, without neglecting them. This is the result of value pluralism which defines the moral and political context of modern democratic societies. It is therefore possible to favour 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 better heal 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 experts' role is not to solve the ethical dilemmas brought on by artificial intelligence themselves, nor become legislators. What are the experts doing then? The experts involved in the *Montreal Declaration* co-construction process don't intend to think for the citizens and suggest a legal and ethical framework that the citizens would merely rubberstamp. Expertise must be at the service of citizen reflection when considering complex social and ethical AI issues.

Sometimes ethicists can give off the impression of looking to preach, of knowing the answers to the touchy questions that the public is asking themselves, and even of being able to preemptively solve tomorrow's problems. It's important to specify their role. Ethicists play three modest but crucial roles:

> They must ensure that favourable conditions are in place for citizen involvement;

> They must clarify the ethical issues that underlie the controversies surrounding artificial intelligence;

> They must rationalize the arguments being defended by the participants by showing them the arguments we know to be wrong or biased and explaining the reasons why they are wrong.

The role of ethicists is therefore that of informed guidance. Experts in other research fields (computer sciences, health, safety, etc.) also play a guidance role by providing participants with the most useful and reliable information regarding the object of controversy (How does an algorithm that learns to make a diagnosis work? Can a doctor be replaced by a robot programmed for the diagnosis? What are the protections we can put into place to defend against attempts to hack our medical data? etc.)

And yet, it must be recognized that the experts themselves often show important cognitive biases. They can be too optimistic or pessimistic towards new technologies they know well; they also tend to put too much weight into their own opinion, especially when they believe they can predict the evolution of their field of society, of social trends, etc. It's by involving them as citizens in the co-construction workshops that we reduce the biases linked to expertize, as well as the authority effect caused by the knowledge imbalance with the other participants.

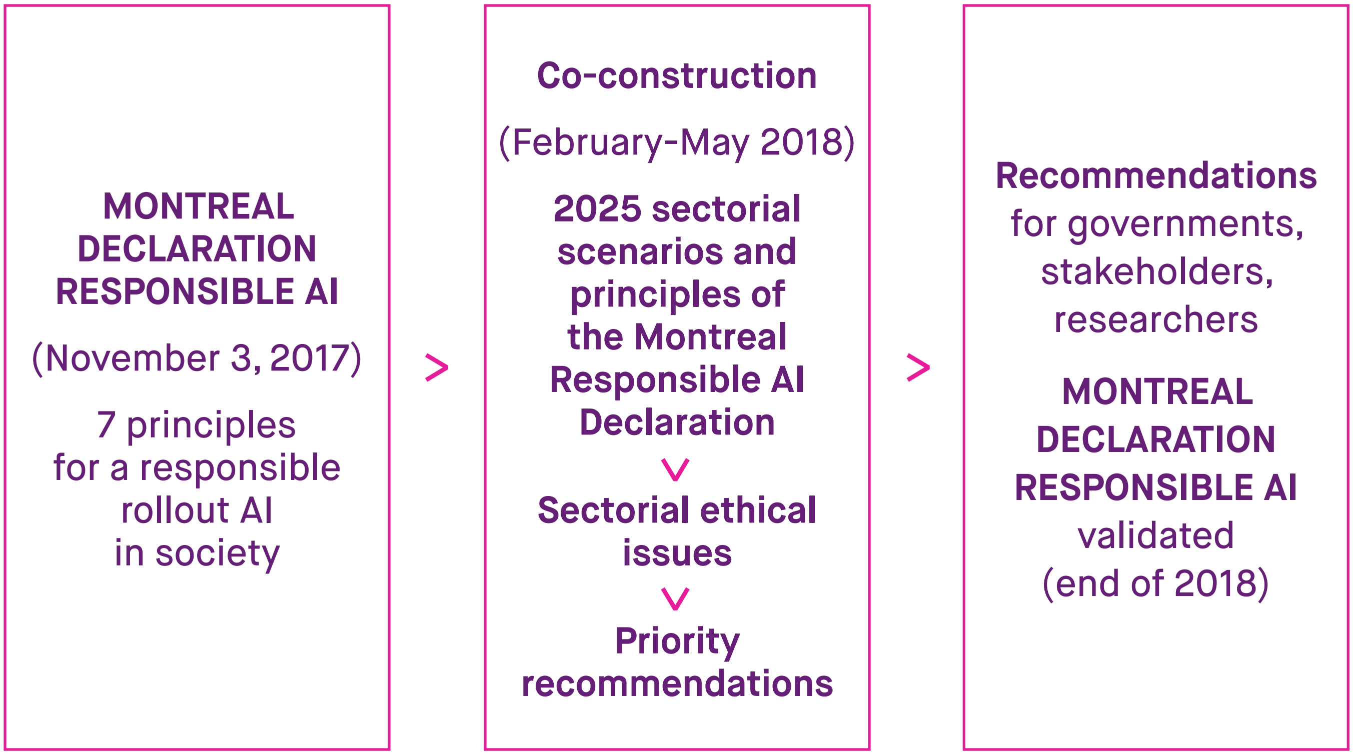
The co-construction workshops are participation spaces that help give direction to the socially desirable development of AI and innovate through proposals that shake up the recognized analysis framework. This essential contribution from citizen deliberations is then analyzed and expanded upon by work committees made up of experts from different fields (researchers, professionals). This work of expanding and drafting recommendations follows the direction defined by the deliberation and remains faithful to the proposals issued at the co-construction workshops.

4.2

THE CO-CONSTRUCTION WORKSHOP METHODOLOGY

The first version of the *Montreal declaration* on Responsible AI, presented November 3, 2017, during the Responsable AI Forum, is the foundation of 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"), it's a matter in this new phase of anticipating with citizens and stakeholders how ethical controversies surrounding AI could arise in the next few years (in the fields of health, law, smart cities, education and culture, the workplace, public services) to then imagine how they could be solved (for example, with a device such as sectorial certification, a new actor mediator, a form or a standard, a public policy or a research program).

The goal of the co-construction approach and its workshops is namely to exemplify and test the principles of the *Montreal Declaration for Responsible AI* thanks to potential scenarios. Ultimately, the process will help specify sectorial ethical issues, and then formulate priority recommendations to the AI community.



MONTREAL DECLARATION RESPONSIBLE AI

(November 3, 2017)

7 principles for a responsible rollout AI in society

>

Co-construction

(February-May 2018)

2025 sectorial scenarios and principles of the Montreal Responsible AI Declaration

Sectorial ethical issues

Priority recommendations

>

Recommendations for governments, stakeholders, researchers

MONTREAL DECLARATION RESPONSIBLE AI validated (end of 2018)

More than ten co-construction workshops took place between February and May: 3-hour citizen cafés in public libraries, and two big co-construction days with various citizens, experts and stakeholders (at the SAT in Montreal, then at the Musée de la civilisation in Quebec City).

The choice of organizing citizen cafés in public libraries is directly tied to the current reinvention dynamic of these public services in Quebec and Canada. By going from a lending space to that of an inclusive "third space" library that seeks to strengthen the capacities of all its citizens (ex. with digital literacy services, citizen support, cultural mediation and discussion areas, the lending of tools and the creation of Fab Labs), public libraries will most certainly have a key role in the responsible deployment of AI in Quebec and Canada[17].

The co-construction days were held in symbolic spaces (Société des arts technologiques in Montréal, Musée de la civilisation in Québec) and namely focused on the meeting between stakeholders and the very diverse disciplines that must work together to imagine a responsible deployment of AI in society.

[17] 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.

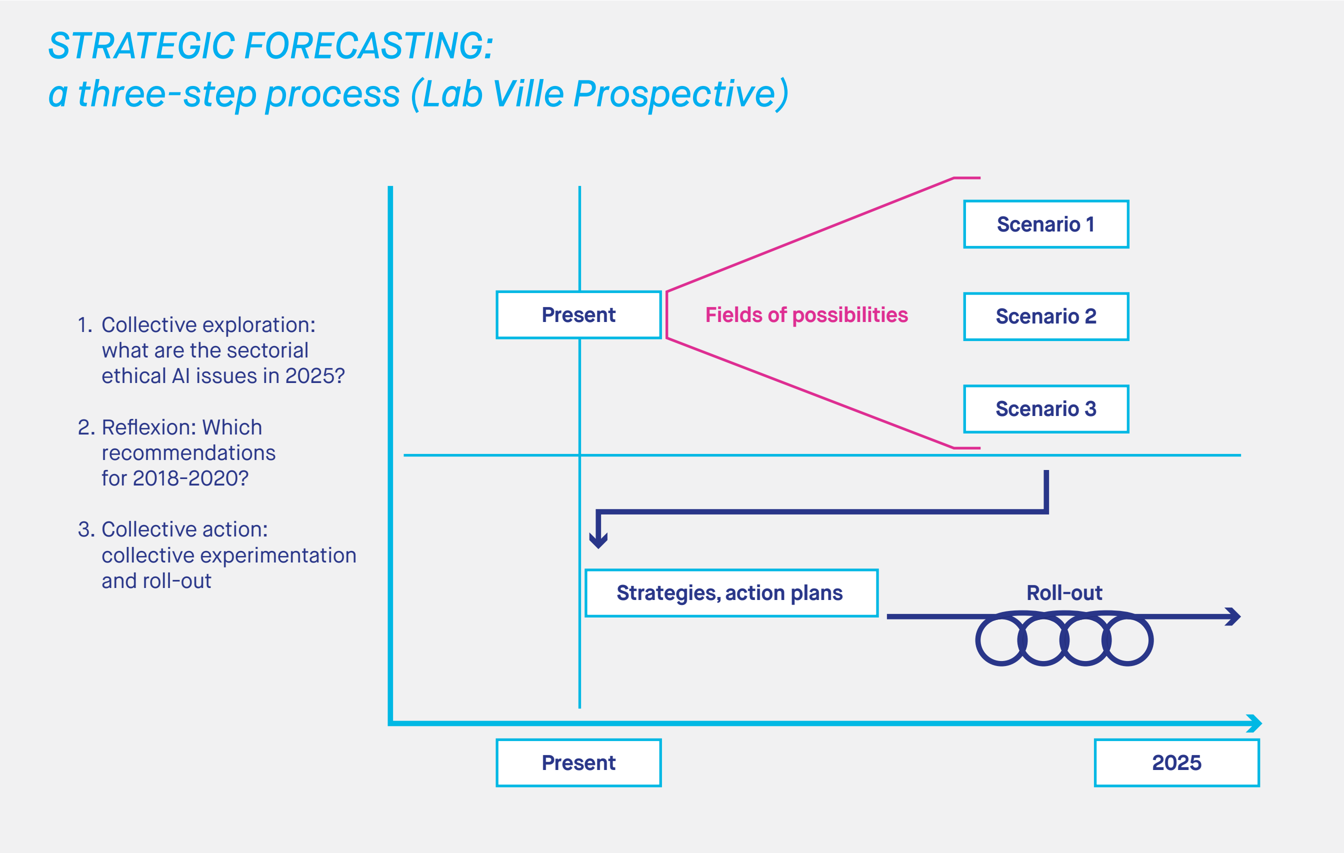
4.3

ORIGINALITY OF THE CO-CONSTRUCTION APPROACH

When compared with other AI ethics initiatives currently underway in the world, this co-construction approach will present four particularly original and innovative dimensions:

> The use of foresight methods, with sectoral scenarios set in 2025 exemplifying through short tales how ethical controversies about AI could appear or increase 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 the face of a wide-open future, will be used to spark the debate, to identify, specify or anticipate sectorial ethical issues on the deployment of AI in the coming years. These discussions with a 2025 horizon can then help retroactively formulate concrete recommendations for 2018–2020, to guide us towards collectively desirable situations.

Figure 3 : Strategic forecasting: a three-step process



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

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

> The use of participative design facilitation methods in multidisciplinary "hybrid forums"[18], including citizens and stakeholders, in a context of shared uncertainty in the face of possible futures (to flesh out a scenario, design ways to respond to an ethical risk, suggest an addition to the Declaration in the case of an orphan issue, i.e. without a corresponding principle).

> Lastly, paying attention to the "paradigm biases" that have very powerful reframing effects in the way they position problems (ex. tackling the ethical issues of self-driving cars strictly from the tramway dilemma angle [ex. MIT's Moral Machine site] and in the context of the "speed-distance" paradigm in transport design), in order to ensure a plurality of issues and draw attention to still unknown or very emerging situations in a rapid change context.

This co-construction approach aims to create a **learning trajectory** to develop, throughout the events, a facilitating kit that's reproducible and user-friendly adaptable and open, that could be published in "open source" at the end of the co-construction approach.

The detail of the world cafés and co-construction days can be found appended to the report.

4.4

WORLD CAFÉS OUTSIDE LIBRARIES

We must also mention the involvement of two philosophy students at Université de Montréal, Pauline Noiseau and Xavier Boileau, who organized many world cafés in spaces other than libraries, and whose formula was more focused on organic discussions about an AI issue. Moderators used very short scenarios, and hosted 2-hour sessions. These sessions were strong deliberation moments with citizens that wanted nothing more than to be involved in public debates, but that were rarely called upon. 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 around AI issues. From an AI scenario on household connected objects (a smart refrigerator), this session namely sparked original reflections on cooking as a relational human activity, raising issues of authenticity, of affection ("a touch of love"), and of social ability, issues that hadn't come up in other consultations based on the same scenario.

[18] Callon, Lacoumes, Barthe, *Agir dans un monde incertain. Essai sur la démocratie technique*, Paris, Le Seuil, 2001

4.5

PORTRAIT OF THE PARTICIPANTS

Recruiting citizens, experts and professionals from different fields of work helped reach a diversity of participants for the co-construction. University faculties, as well as inter-university research centres and their networks helped reach an important number of players involved in the development and use of AI.

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

Notable fact, there was practically the same number of men and women in all workshops. A strong majority of participants had post-secondary education and were in the 19-34 age group.

Figure 4 : Proportion of men and women involved in the co-construction workshops

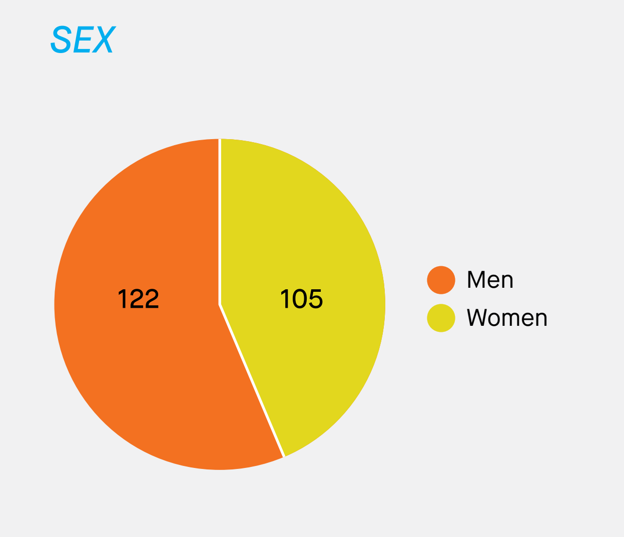


Figure 5: The participants in the co-construction workshops by age groups

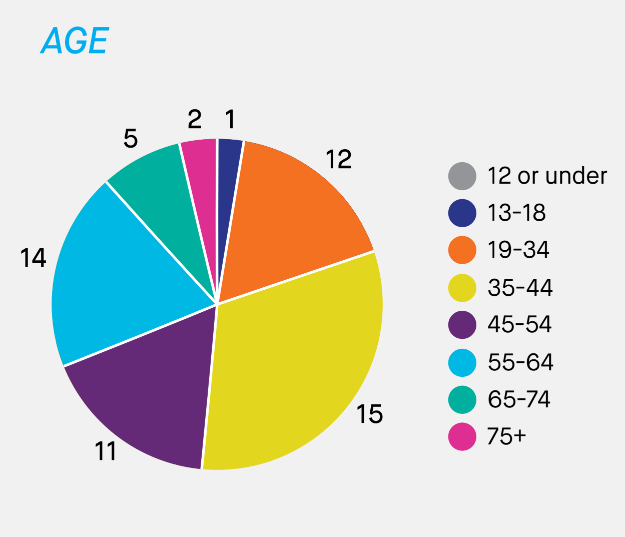


Figure 6 : Distribution of participants in world cafés and co-construction days by education level reached

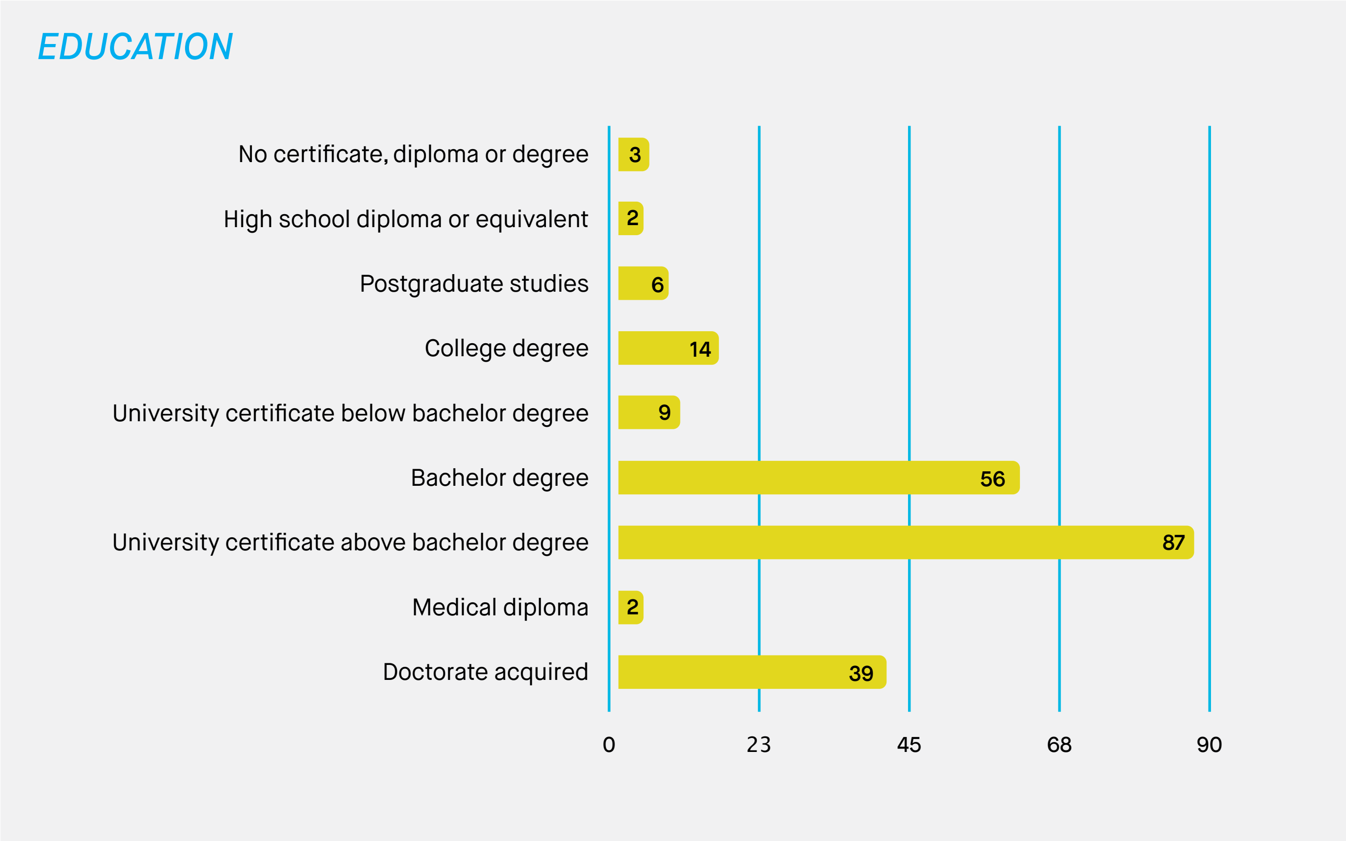
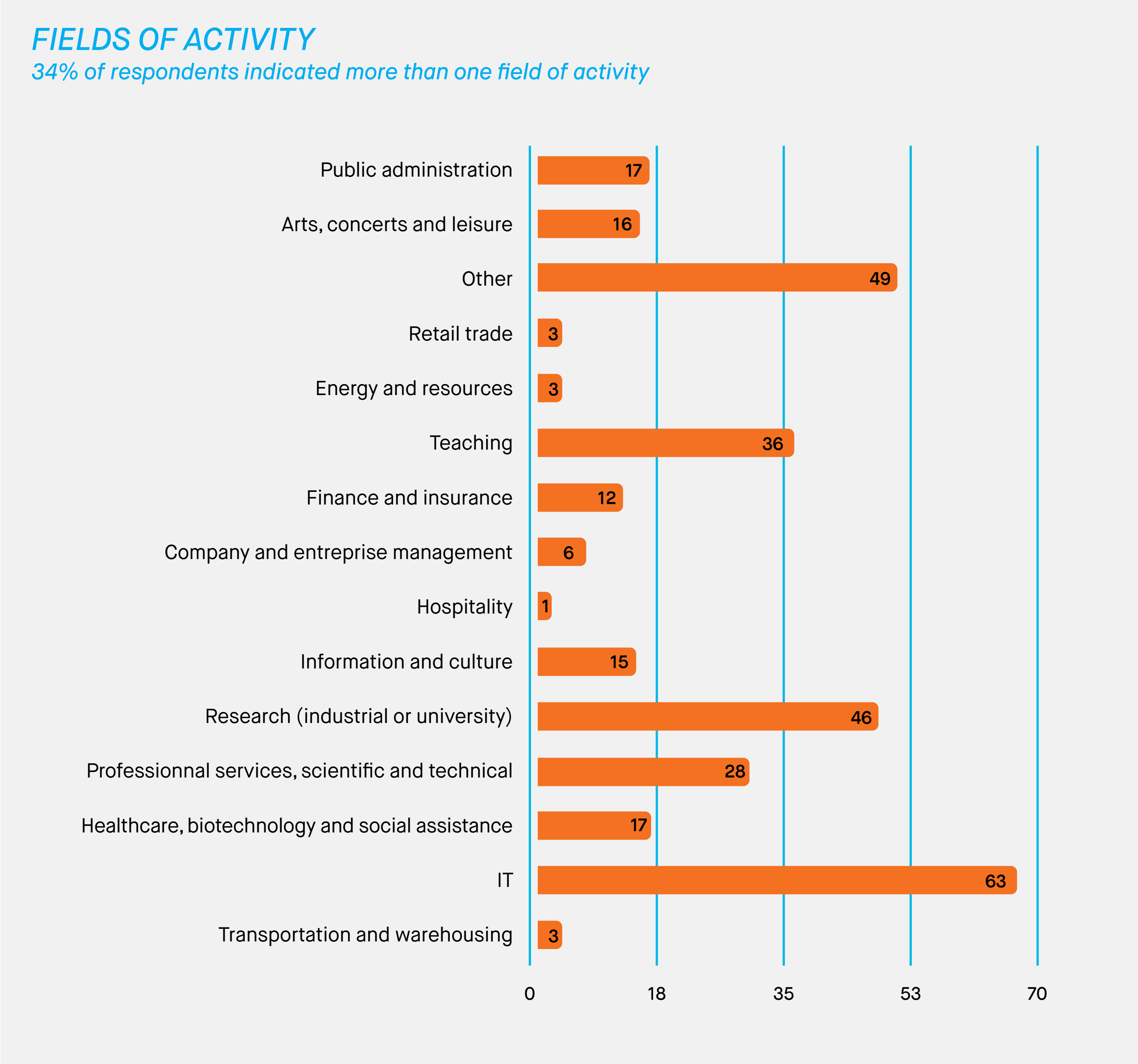


Figure 7: Distribution of participants in world cafés and co-construction days by field of activity



FIELDS OF ACTIVITY

34% of respondents indicated more than one field of activity

5. DELIBERATION PATHS IN THE WORKSHOPS

EXAMPLES FROM TWO ELEMENTS: SMART CITIES AND THE WORKPLACE

5.1

THE DELIBERATION PATHS

How did the discussions and deliberation in the co-construction workshops unfold? What kinds of reactions did they provoke? What were the main points of discussion that led to recommendations for an AI framework? This section of the document details certain 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 paths of two teams representing two sectors among the five discussed in the co-construction were selected:

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

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

First step: formulating **sectorial ethical and social issues in 2025** (by cross-referencing the general principles of the *Montreal Declaration* with the 2025 user situations described in the debate-provoking scenarios): the formulating of individual issues (on Post-its) was then expanded upon in a collective discussion from which came a selection of three priorities.

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

Third step: he storytelling of the launch of a first recommendation in 2020 (the newspaper headline) to take stock of the **"time for collective action"** with its organizational constraints: from formulating ideas to synthesizing them in orderly fashion within a narrative.

We must mention that between these steps and micro-steps of the deliberative path, the "nature" of the ideas generated varies: some are individual intuitions (when, at the start of the exercise, participants write down many sectorial issues on Post-its), others stem from a collective discussion (where each person justifies their point of view), and others yet are the result of a hierarchy determined by the group (when selecting three key issues to write on the summary poster).

We therefore find in these prospective workshops three properties of the deliberative devices highlighted by Blondiaux and Sintomer in their article *L'impératif délibératif*[19] (Politix, 2002, pp. 25-26): allow the imagination of new solutions in an uncertain world; allow a rise in generality and aim for consensus or "deliberative disagreements" in a society marked by the pluralism of values; and finally, provide a factual and normative source of legitimacy through the inclusion of everyone in these deliberations.

[19] Blondiaux L. et Sintomer Y., « L'impératif délibératif », *Politix*, 2002, p. 25-26.

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 SDC are circulating in Montreal and controversy arises around sharing the road and public spaces. Some lanes are now reserved for SDC and are protected by barriers, so that they can drive at a moderate, but fluid speed (50 km/h) without risking accidents. SDC can also drive elsewhere, but at very slow speeds (25 km/h). Protesters for active mobility (walking, biking) disturb the operation of these protected lanes, knowing that the algorithms of the SDC are set to "altruistic" mode to protect outside people.

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

The deliberative path presented is the result of a 3h table in a Montreal public library, with eight citizens interested in new technologies and otherwise making active family mobility habits (walking, biking). From this scenario set in 2025, the discussion led to formulating an initiative presented as a headline in the March 13, 2020, edition of the Responsible AI Gazette: "First autonomous mobility literacy workshop held." What was the deliberative path that led to this original proposition? What were the defining moments? How did the ideas grow at every step? We present and comment certain significant moments of the path taken by this team.

FIRST DELIBERATIVE MOMENT: FORMULATING ETHICAL ISSUES IN 2025

Many interrogations drawn up on Post-its were submitted by the participants in relation to different principles of the *Montreal Declaration*:

THE AUTONOMY PRINCIPLE

"Will humans become too dependent when it comes to moving?", "Will the freedom of movement be impeded by AI?", "We're giving up a lot of micro-decisions to AI and interconnected systems, at the expense of humans."

THE WELL-BEING PRINCIPLE

"A lot less room for spontaneity with SDC", "What will the neighbourhood development look like in regards to the road axis of SDC?", "Will transportation data influence city urbanization?"

THE DEMOCRACY AND JUSTICE PRINCIPLE

"What is the difference in installing transportation axis in working-class neighbourhoods as opposed to affluent neighbourhoods?", "Will only those who are well-located get to enjoy the fluidity of traffic?"

THE PRINCIPLE OF PRIVACY

"Will we be able to replace anyone's movements?" of responsibility: "Who will be responsible for an accident?" or of security: "Possibility of hacking fleets of vehicles?" this last principle coming from the participants, in addition to those found in the declaration.

Many in-depth discussions then occurred, participants bouncing off of the first ideas to generate new ones on spontaneity and freedom to travel, on the safety of personal data and its management by a central organization, on the question of algorithm settings and the possibility of manipulating them.

Then, after a nearly 45 minute-long discussion, the participants used coloured stickers to select 2025 ethical issue groupings that seemed a priority to them. The participants voting with coloured stickers on the wall with Post-its and discussed ideas associated with four principles of the *Montreal Declaration*, two of which were regrouped: safety, justice, and well-being and autonomy.

Table 1: Smart City, First deliberative moment: formulating ethical issues in 2025

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

This selection of priority issues by the team is an original one: although the issues of security, responsibility and privacy are often raised in studies and debates on SDC, those of justice, well-being and autonomy are much less discussed.

SECOND DELIBERATIVE MOMENT: **RECOMMENDATIONS FOR AN AI FRAMEWORK IN 2018-2020**

To respond to these issues, the team chose to continue its discussions by trying to think about the four associated principles. Many AI framework recommendations were formulated by the participants. We present three (out of six) here, which allow you to follow the path of an idea all the way to the headline of a newspaper.

Table 2: Smart City, Second deliberative moment: AI framework recommendations for 2018-2020

Framework recommendations for 2018-2020	**1**	**2**	**3**
Description	Training for collective vigilance (ex. driver's licence)	An all-party committee that manages incidents, injustice and other issues in democratic fashion; the committee must be a decision-making one	Evaluating the urbanization plan during the transition period
Instrument Categories	New training	New institutional player	Participative planning process

These recommendations, which show true institutional creativity (beyond the very broad examples of tools provided in the participant booklet), are in line with the issues identified at the previous step, but also present an enrichment of ideas (they are not simple deductions from tools adapted from an identified ethical case).

The idea of training for vigilance and of participating in a collective decision (through an all-party committee and open planning) do indeed lead to recommendations for capacity building and local forms of democracy.

THIRD DELIBERATIVE MOMENT: **WRITING A HEADLINE FOR A 2020 NEWSPAPER**

These measures were then storyboarded in the following way 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 established 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; sharing the rod between pedestrians, bicycles, SDCs, trucks; explaining the rules; trial sessions; the question of algorithm settings."

This newspaper headline, which was formulated after a discussion among the participants, contributes again to the progression of ideas. Indeed, the principle of a workshop on **"autonomous mobility literacy" allows the creation of new meaning** by integrating the various recommendations formulated in the previous step widening the scope to take about autonomous mobility and not simply SDCs (thus allowing for the possibility of autonomous multimodal transportation). This headline also presents a **collective action device** with a progress target (the training and abilities of citizens, the possibility to participate in the city's decision-making committees regarding SDC rollout) and an organization (a rollout in public libraries across Quebec, which are currently transforming into cultural services third parties for all citizens.

The result of this table is particularly interesting because it helps consider the ethical question of self-driving vehicles from the perspective of autonomy and social justice in the city, and not strictly from a responsibility in case of accident scenario, as MIT's Moral Machine initiative does, for example, from the moral dilemma of the tramway[20].

[20] MIT site: 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 in order 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 this restructuring in socially responsible fashion, for instance by creating a data processing coop 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 issues concerning the change in the process caused by AI that thousands of SMEs and big businesses in Quebec will be faced with between 2020-2030.

The deliberation path presented in this section comes from a table eld over an entire day in Montreal bringing together nearly ten researchers and experts working on workplace mutations, social and the social responsibility of businesses and unions. A citizen that had previously attended a workshop in a public library was also at this table.

Starting from the 2025 scenario, this team's work led to the formulating of an initiative that made the headline of the February 18, 2020, Responsible AI Gazette: **"First measures of the mixed interdepartmental committee on responsible digital transition."** Like in the previous case about self-driving cars, what was the deliberative path that led to this original proposition? What were the defining moments? How did the ideas grow at every step? We present and comment certain significant moments of the path taken by this team.

FIRST DELIBERATIVE MOMENT: **FORMULATING ETHICAL ISSUES IN 2025**

Many Post-its were drafted by the participants in the first part of the morning workshop. Here are a few of them and an overview through a few formulas taken from Post-its and the table of grouping by *Montreal Declaration* principles.

Certain formulated issues were associated with different *Montreal Declaration* principles:

THE WELL-BEING PRINCIPLE

"What do we favour? The company or society?", "Adopting different perspectives on well-being: individual (employee), social and collective development, economic development (SME)", "What do performance ideals look like when robots or co-bots never get tired, unlike humans?", "What are the possible positive aspects: professional reinforcement, for ex. in medicine, less drudgery for certain positions", "What are the new forms of work and protection with work/leisure?"

AUTONOMY PRINCIPLE

"What professional and life paths? Can you choose not to change careers because of AI? What are the consequences?", "collective autonomy: for the collective and critical anticipation of discussion on the urgency of adaptation"

THE RESPONSIBILITY PRINCIPLE

"Who is held responsible for these changes?", "Is the social and ethical responsibility of the transition individual — each company – or collective —society, the government?", "What funding for the transition?"; "How to align the cost-effective directive and the responsibility in an emergency context?"

THE KNOWLEDGE PRINCIPLE

"What collaboration between humans and robots? Workload, health and safety, training, acceptability, cybersecurity," "How is data collected in a context where this type of work is mainly carried out by private corporations (GAFAM)?", "How to prevent people getting stuck in classes?", "What are the possibilities of data being shared?", "What is the impact on the educational system?"

THE JUSTICE PRINCIPLE

"What independence in the face of power being concentrated among GAFAM?", "What social redistribution of the social benefits of AI?", "Will the productivity gains created by AI and industry 4.0 be sufficient to fund the social transition if companies practise tax evasion?", "What equity in case of sharing and coding an employee's implicit knowledge to transform it into data or feed the automation?", "Do we have a choice, as employees, not to reveal this information?", "On what criteria will we choose those who are replaced and those who are trained?", "What access to the social protection of tomorrow?", "What access to rights, such as the right of association, in this new workplace reorganization?"

THE DEMOCRACY PRINCIPLE

"Is precariousness a fatality when the transition can be anticipated?", "the politicized short-term vision rather than a long-term vision", "the obscuring of decision-making processes", "risks of biases in the algorithm training sets", "the need for a democratic debate".

We can mention here that the typology of the *Montreal Declaration* on responsible AI principles worked well to provide benchmarks for the discussion, and that the participants even suggested original problems concerning certain principles: the necessity of addressing **well-being** and **responsibility for the transition** from different points of view (individual and collective); the relationship with social time, with an opposition between the collective anticipation and the opaque language of urgency, as a condition of our **collective autonomy** and exercising our **democracy** (the lack of time preventing well-informed democratic work); a strong requirement for **justice** in the social redistribution of AI benefits, namely in terms of equity accompanying the codification, and therefore possible automation, of employee skill sets.

After a good hour of discussion, the participants used coloured stickers to select groupings of 2025 ethical issues they deemed priorities. The votes being spread out pretty evenly over the various issues, all deemed equally important by the group, the formulation of three priorities for the poster was also a synthesis exercise of the ideas discussed in the first part of the workshop (see table below).

Table 3: Workplace, First deliberative moment: formulating ethical issues in 2025

2025 Ethical Issues	**1**	**2**	**3**
Description	**Heavy concentration of power** (see GAFAM) that prevents: - An equitable sharing of AI benefits - the arrival of new players (new COOP type business models) - reduce inequities (literacy)	**Technological determinism, fatality ("Black box society") and urgency:** instead of taking the time to have an informed, participative, democratic debate on new social risks, social development models, performance ideals, work experience.	**Defining the common good and the type of collective responsibility in the digital transition** For example: which stakeholders? The company alone? The State? Unions? The educational system?
Associated Principles	Justice and independence	Democracy, knowledge and collective autonomy	Well-being and responsibility

SECOND DELIBERATIVE MOMENT: **RECOMMENDATIONS FOR AN AI FRAMEWORK IN 2018-2020**

To respond to these issues, the team continued talks in the afternoon by going around the table once more leading to the drafting of AI framework recommendations by the participants, which then led to numerous recommendations that were collectively discussed one by one. The table below presents an excerpt (six propositions out of the more than 10 that were formulated by the group), in order to follow the path of an idea up to the formulation of a newspaper headline.

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

2018-2020 framework recom-mendations	**1**	**2**	**3**	**4**	**5**	**6**
Description	**Reinforce digital literacy for all.** With a skill set reference for public libraries, schools, and the workplace. By dealing with the question of illiteracy and "non-use" by citizens	**Mixed permanent interdepart-mental committee on AI, executive next to PM.** At the interface of the economy, employment, education and culture themes (see digital strategy)	**Digital AI insurance funds** to enable training and adaptation. Example of device: the 50-week Parental Insurance Plan which can also inspire a minimum income to prevent pre-cariousness.	**Incentives on new business models for data processing** Example: COOP model to break the isolation of self-employed workers treating data and ensure collective autonomy	**Guiding investments towards responsible AI for the common good** SRI (Socially responsible investment) model. Investments from the State, from individuals in synergy with the worker's fund	**Accelerated process to update and create professional programs** With cégeps, universities, departments, professional orders impacted by AI (ex. law, healthcare)
Instrument Categories	New training	New institutional player	New insurance mechanism	Incentive	Funding device	Planning process

As in the previous case about self-driving cars, these recommendations, which show true institutional creativity (beyond the very broad examples of tools provided in the participant booklet), are in line with the issues identified at the previous step, but also present an enrichment of ideas. If digital literacy is indeed a goal in the policy's agenda (ex. Stratégie numérique du Québec), it's the necessity that it expands that was highlighted. The other recommended measures are innovative and invite the creation of a new public, all-party or collective devices to ensure Quebec society's true autonomy when faced with AI issues in the workplace. In that sense, the group has chosen collective responsibility towards AI in its transition into society.

THIRD DELIBERATIVE MOMENT: **WRITING A HEADLINE FOR A 2020 NEWSPAPER**

These measures were then storified for the poster. The headline formulated by the team for the February 18, 2020, Responsible AI Gazette reads as follows:

First measures of the mixed interdepartmental committee on responsible digital transition

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

1. A new digital insurance fund worth 2 billion dollars (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 cooperatives (against precariousness).
4. A literacy fund worth 10 billion dollars over 5 years, on the basis of a new skill set inventory.

This newspaper headline, which was formulated after a discussion among the participants, contributes again to the progression of ideas. Indeed, the mixed interdepartmental committee on responsible digital transition would be a creation. This new institutional player, born from a reflection on a 2025 scenario concerning the impact of AI on the Quebec workplace, could represent a new common step for many public policies that successfully address the digital transition and the issue of digital literacy, but don't raise 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 player, which could be 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 created by the rollout of AI in Quebec's public and private organizations.

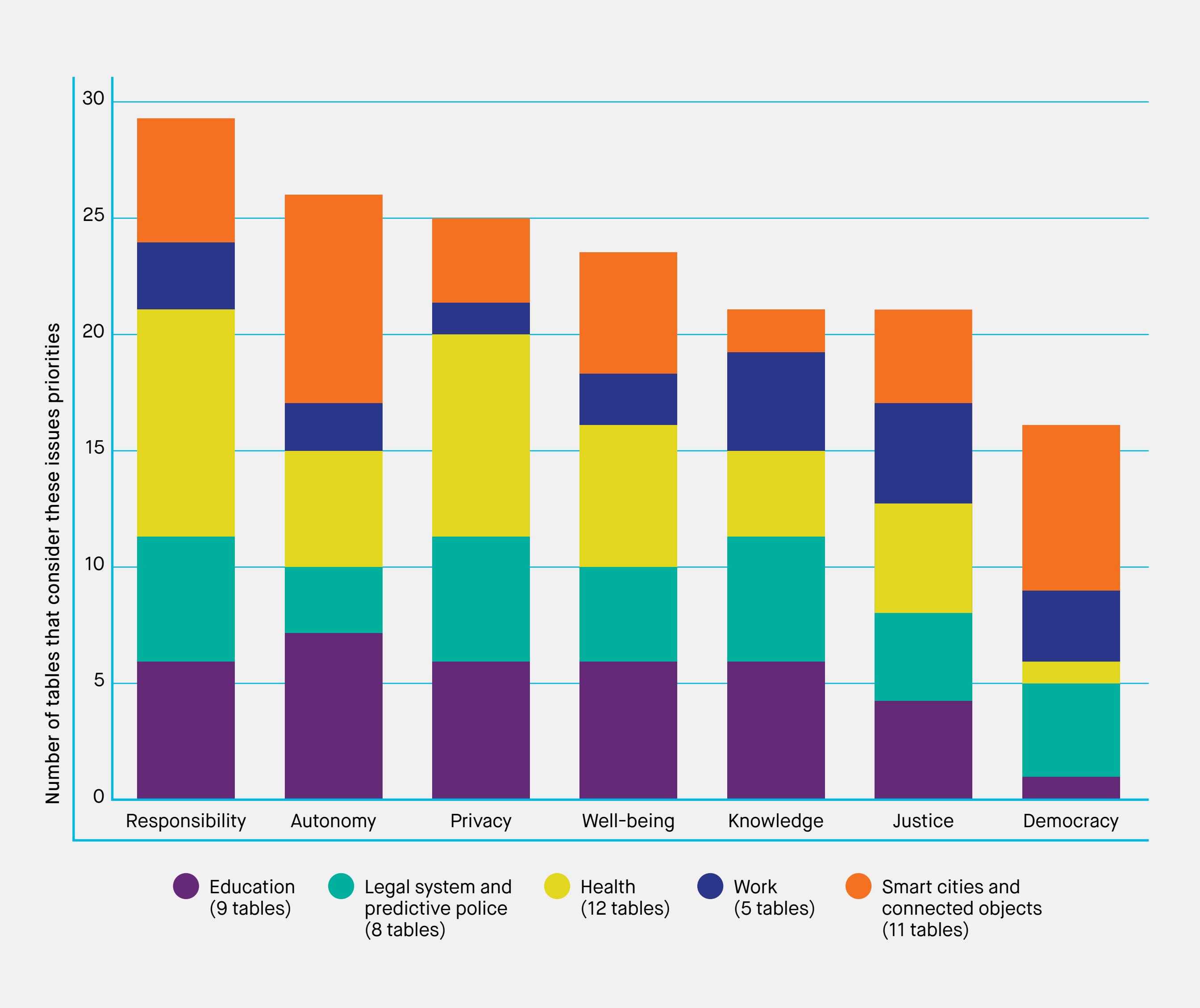
6. THE FIRST CO-CONSTRUCTION RESULTS

6.1

RÉSUMÉ

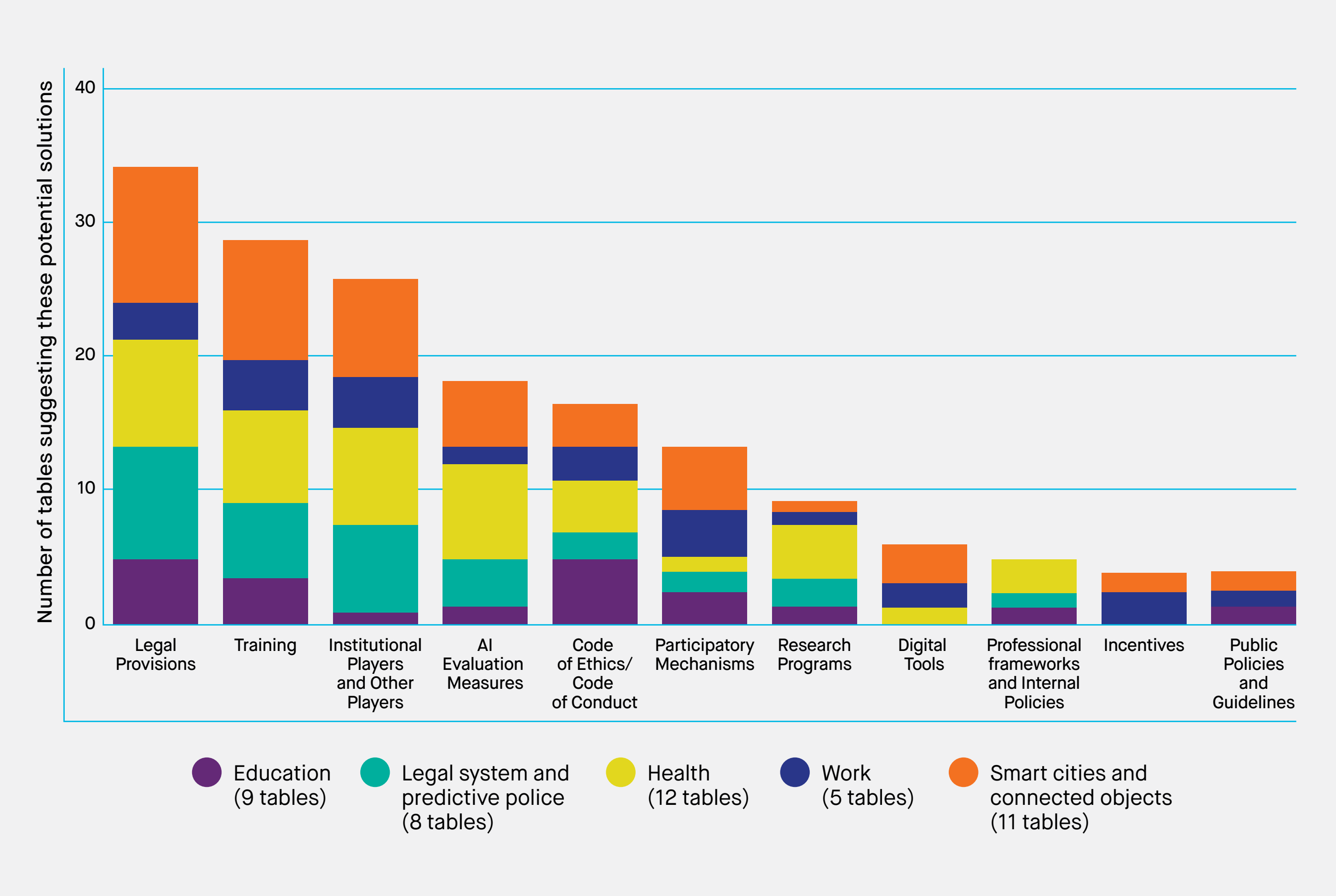
Citizens gathered around 45 tables to discuss their perception of issues tied to applying the Declaration's principles.

Table 5: The principles that the priorities identified by citizens refer to



They identified different potential solution categories to respond to these issues.

Table 6 :Suggested potential solutions to respond to the identified issues



6.2

CO-CONSTRUCTION DATA: EXPLANATORY NOTES

The current section relates the results collected during the co-construction tables held in winter 2018 for the *Montreal Declaration*, 45 tables in all that brought together hundreds of citizens. Discussions were held around 5 major sectors of AI development: the education sector (9 tables); the justice and predictive police sector (8 tables); the healthcare sector (12 tables), the workplace sector (5 tables), and the smart city and connected object sector (11 tables).

These results stem from a preliminary and non-exhaustive analysis of the two main axis discussed at each table: the various **issues** raised by AI development, as well as the **potential solutions** identified that came out of citizen discussions based on provoking scenarios. At this level, the analysis remains descriptive and as close as possible to the citizen's words. For the purposes of this report, emphasis was placed on: 1) the great directions expected in terms of responsible AI development; 2) the presentation of issues citizens determined to be priorities; 3) the issues that could lead to the creation of new principles in the Declaration; 4) the potential solutions identified by citizens to respond to these issues.

The great directions expected in terms of responsible AI development refers to citizen recommendations that are not specified in concrete potential solutions. They nonetheless allow the main positions and standard expectations citizens have towards AI development.

Each co-construction table was invited to choose 2 or 3 issues to be treated as priorities before 2025. Only issues that were considered priorities by citizens were analyzed for the purposes of this report. These priority issues were therefore described on the basis of citizen formulations and classified, for each sector, according to the Declaration principles they are linked to. However, it's worth noting that just because certain issues weren't considered priorities that they weren't discussed, that they're less important, or that the principles weren't discussed for each sector. One single principle for each sector is detailed in this progress report.

Different issues that could lead to the creation of new principles in the Declaration were identified on the general basis of the discussions that took place. In this report we present, in non-exhaustive fashion, those that proved to be particularly relevant.

Finally, the potential solutions identified by citizens to respond to these issues have been classed in 11 main categories. These categories will be specified in subsequent steps of the analysis. The category that seems most relevant to bring up for each sector is presented in greater detail.

Regarding the quantitative data in this report, the number of occurrences corresponds with the number of tables where each issue/potential solution was formulated in consensual fashion, in conformity with the co-construction process.

The total number of potential solutions (n=190) corresponds to those identified as priorities by citizens (since they were invited to clearly formulate them on posters). However, potential solutions mentioned during the discussions but not explicitly appearing on the posters are also taken into consideration.

Quotes from the report are presented in such a way as to reference the co-construction table when they come from a group formulation (consensus). Other quotes correspond to individual formulations (written on Post-its by participants or copied verbatim by members of the group).

6.3

RESPONSIBLE DEVELOPMENT OF AI: THE GREAT DIRECTIONS EXPECTED BY CITIZENS

Generally speaking, the participants recognized that the arrival of AI came with important potential benefits. Namely, when it came to work and legal matters, participants recognized the time savings that AI devices could bring:

"It would help reduce wait times to treat cases."
- A participant

However, it was also mentioned that AI development had to be done with caution and right now to prevent abuse, although some consider the possibilities brought on by AI to still be limited.

The implementation of a framework is therefore recognized as necessary to prevent risks rather than determining who is to blame when they occur:

"You don't care so much about knowing who to sue when things go wrong, you want to find ways to make sure things don't go wrong in the first place."
- A participant

The citizens highlighted the need to implement different mechanisms to ensure the quality, intelligibility, transparency and relevance of the information being communicated. They also discussed the difficulty of guaranteeing truly enlightened consent.

The great majority of the participants recognized the necessity to align public interests with private ones and prevent the apparition of monopolies, or limit the influence of corporations (which are seen as ungovernable) through more cohesive and legal measures. These mechanisms should be, as much as possible, simple and changing so they can adapt to the rhythm of AI development and allow its steady control. In the legal sector, certain participants mentioned a "gap" separating technology (defined as quick, innovative, even abstract) and our institutions (often too stiff in their integration of technology) that aren't able to deal with these changes in society. Some tables went as far as suggesting "nationalizing AI", which would then "become a public service, and programmers would be public servants". (Smart City and connected objects table, INM, Montreal, February 18, 2018, Connected refrigerator scenario).

The participants also recommended guaranteeing a contextual approach to AI, which must take different parameters into account (ex. mandatory or optional collection of data the algorithm learns from). These mechanisms should come from and involve independent, trained people to favour the diversity and integration of those who are most vulnerable, and protect the mixed aspect of the lifestyles.

Whatever the use, the majority of the participants insisted on the fact that AI must remain a tool, and that the final decision must come from a human being (whether it's a legal ruling, a hiring decision or a health diagnosis), which implies recognizing its limitations.

"AI proposes, mankind disposes."
- A participant

The protection of an individual's privacy and the management of personal data were heavily discussed. For example, processing healthcare data should be managed in a special way, given the highly sensitive nature of the information. It should therefore both favour methods of control ranked according to the type of use and adopt security as an operational mode. Regarding the workplace sector, the participants recommended the obligation to inform users of how their data is processed.

Aware that these recommendations involve important institutional changes, participants highlighted the need to keep in mind that AI is not necessarily desirable to begin with.

"Just because you can, doesn't mean you should."
- A participant

The citizens generally agreed that the consequences of AI use in the different sectors—for both the individual and society as a whole — must clearly be measured to establish benchmarks without unduly hindering progress.

6.4

RESULTS: CITIZEN PERCEPTION OF THE ISSUES

The great issues of responsible AI development

Table 7: Priorities identified by citizens according to Declaration principles (number of tables)

	Education	Legal system and predictive police	Workplace	Healthcare	Smart city and connected objects	Total number of tables that consider these issues to be priorities
Responsibility	6	5	3	10	5	29
Autonomy	7	3	2	5	9	26
Privacy	6	5	1	9	4	25
Well-being	6	4	2	6	5	23
Knowledge	6	5	4	4	2	21
Justice	6	4	5	4	4	21
Democracy	1	4	3	1	7	16
Total number of co-construction tables	9	8	5	12	11	45

Citizens that took part in the co-construction days were invited to select 2 or 3 issues to address as priorities before 2025 regarding the responsible development of artificial intelligence.

The responsibility principle was the one most often deemed a priority, followed by the autonomy principle, privacy, then well-being (individual and collective), knowledge and justice. It's worth noting, however, that they are all closely intertwined.

The principles of knowledge, responsibility, privacy, justice and democracy are presented below by sector. As for the autonomy principle, often selected as a priority, it concerns the preservation, even encouragement of individual autonomy opposite risks of technological determinism and reliance on tools. It also raises the issue of double freedom of choice: being able to make your own choice when faced with a decision guided by AI, but also being able to choose not to use these tools without risking social exclusion. The freedom included in this autonomy principle regarding AI systems would involve any person's capacity for self-determination.

"Develop technologies that favour human autonomy and freedom of choice."

(Education table, Bibliothèque de Laval, March 24, 2018, Hyperpersonnalisation of education scenario).

The well-being principle also holds an important place for participants. It's there, pervasive at every table, exhibiting a collective desire to move towards a society that is fair, equitable and that favours the development of all. Well-being is therefore both a collective (tied to equity and accessibility issues comprised in the justice principle) and an individual issue, aiming for the fulfillment without impeding on autonomy and privacy. Participants showed a preference for AI development "that would allow any individual to access personal and social fulfillment". (Education table, Bibliothèque Père Ambroise, Montréal, March 3,2018, AlterEgo scenario).

Broadly speaking, the well-being principle was also behind a call to maintain quality human and emotional relationships between experts and users in all fields.

6.4.1

KEY ISSUES BY SECTORS

EDUCATION

As for the education sector, issues regarding privacy, responsibility, well-being and knowledge were considered priorities by 6 tables out of 9. Discussions about issues dealing with the knowledge principle were especially relevant to introduce the question of transforming human skill sets in an AI era:

ISSUES DEALING WITH THE KNOWLEDGE PRINCIPLE
(6 tables sur 9)

The issues dealing with the knowledge principle for the education theme stem from skill transformation issues in a context where both the role of a teacher and the methods of developing and accessing knowledge are rapidly changing. This principle was mostly discussed under the optics of transforming the learning relationship, which would then become an issue of a teacher's expertise whose work would have to be modified. It was also mentioned in relation to the diversity principle to discuss the need to foster a variety of intelligence and relationships to knowledge.

"Redefining/transforming the nature of the relationship between teachers and students in the classroom and modifying relationships to knowledge."

(SAT Table, Montreal, March 13, 2018, Nao scenario).

"Human skills and abilities: importance of developing many learning environments."

(Musée de la civilisation Table, Quebec City, April 6, 2018, AlterEgo scenario).

LEGAL SYSTEM AND PREDICTIVE POLICE

As for the justice and predictive police sector, issues regarding privacy, responsibility and knowledge principles were considered priorities by 5 tables out of 8. Discussions about the issues concerning the responsibility principle allowed us to clarify the principle's scope:

ISSUES CONCERNING THE RESPONSIBILITY PRINCIPLE (5 tables out of 8)

The responsibility principle was formulated in two main ways: as a demand for human accountability in legal rulings, and by concern for who is responsible for the decision (and any potential error). The algorithm's lack of transparency goes against accountability, in the eyes of the citizens, since it's hard to retrace what is considered in the decision. The responsibility principle is therefore tied to the knowledge and transparency principles in regards to the claim to make decisions explainable and preserving a place for human players and their skill sets in the legal system.

"[Justice] must remain a tool whose sole purpose is to protect individuals. Promoting compassionate and equitable justice that accounts for singularities and past experiences. Artificial intelligence must not have the right to judge human behaviour. The final decision must always require human intervention."

(SAT Table, Montreal, March 13, 2018,
Preventive arrest scenario).

"Transparency, accountability and responsibility in regards to creating the tool, to the data being used, and to the tool's consequences."

(SAT Table, Montreal, March 13, 2018,
Conditional release scenario).

HEALTHCARE

In regards to healthcare, the issues concerning the privacy and responsibility principles were considered priorities, by 9 and 10 tables out of 12, respectively. The issues concerning privacy hold particular significance for the sector given the relatively sensitive quality and near-always personal character of health data.

ISSUES CONCERNING THE PRIVACY PRINCIPLE
(9 tables out of 10)

Participants identified different issues related to confidentiality and invasion of privacy. These issues concern a potential invasion of privacy that can be linked to the development and configuration of AI systems (ex. which should help avoid pirating, shortages and abuse). They also deal with what the citizens called "rétroaction" (use of data previously collected for another purpose) and accessing this data through private companies. Faced with these issues, citizens worried about how to make sure the data isn't sold, and how to guarantee that the patient keeps control over their data (especially when it's private data), and hold imperative rights to them.

"How far are we willing to share our personal data (information) as individuals in order to feed healthcare services?"

(Musée de la civilisation Table, Quebec City,
April 6, 2018, Digital Twins scenario).

WORKPLACE

As for the workplace sector, the issues concerning the justice and knowledge principles were considered priorities (respectively 5 and 4 tables out 5). All the tables that gathered around the development of AI in the workplace therefore considered that the issues concerning justice, equity and diversity should be addressed separately.

ISSUES CONCERNING THE JUSTICE PRINCIPLE
(5 tables out of 5)

The justice principle raises two main concerns: ensuring an equitable sharing of AI benefits among all players, social groups and territories, and *"installing nondiscriminatory algorithms that favour diversity, inclusion and social justice"*. (Musée de la civilisation Table, Quebec City, April 6, 2018, AI as mandatory path to the workforce scenario).

"Sharing AI benefits (productivity gains); equity among social groups, territories (cities and regions), taking vulnerabilities into consideration; the meaning of work in society and in the construction of our identities."

(Musée de la civilisation Table, Quebec City, April 6, 2018, A socially responsible restructuration scenario).

SMART CITY AND CONNECTED OBJECTS

As for the smart city and connected objects sector, the issues concerning the autonomy and democracy principles were considered priorities by 9 and 7 tables out of 11. Many issues seemed to potentially infringe on the democracy principle according to citizens:

ISSUES CONCERNING THE DEMOCRACY PRINCIPLE
(7 tables out of 11)

Participants discussed issues tied to the balance between collective interests and individual needs; to managing access to public spaces and sharing those spaces, or even sharing the benefits stemming from the development of AI technologies (namely, between individuals, the public sector and the private sector). They insisted on the necessity and the difficulty of ensuring a collective (involving citizens) and enlightened (which implies a certain transparency regarding the development of AI systems) decision-making process to define guidelines around connected objects. Citizens also questioned the true independence of public authorities in regards to AI development, and out forward the risk of normalizing behaviour that could lead to marginalization, thereby running the risk of infringing on the democracy principle.

"How can we manage an intelligent transportation system in democratic fashion?"

(Du Boisé Library Table, Montreal, March 17, 2018, Self-driving car scenario).

6.4.2

ISSUES THAT COULD LEAD TO THE CREATION OF NEW PRINCIPLES

Different issues identified or discussed by citizens seem particularly interesting and could lead to the eventual creation of new principles in the *Montreal Declaration*, namely for their transversal aspect (for both the sectors and the principles).

ENVIRONMENTAL ISSUES

For example, the impact of the responsible development and use of AI on the **environment**. These issues ask how to guarantee the responsible and equitable use of material and natural resources. They also raise the matter of ensuring a positive energy balance when it comes to the polluting effects of AI and the technologies associated with its use.

"We forgot to talk about the environmental aspect: the stocking of data, the problem of an outrageous accumulation of data and the costs in terms of energy (or room) that involves."
- A participant

SPECIFY THE JUSTICE PRINCIPLE: DIVERSITY AND EQUITY

The justice principle was discussed according two types of issues: sometimes in terms of diversity, sometimes in terms of equity and social justice. This principle could therefore be split in two to put forward each of these aspects:

> A diversity principle could therefore aim to prevent discrimination by finding mechanisms free of biases tied to sex, age, mental and physical capacity, sexual orientation, social and ethnic origins and religious beliefs, without creating any new ones. The diversity principle also calls upon favouring a multitude of perspectives and intelligence rather than standardizing individual profiles according to a limited number of categories and criteria.

"A loss of diversity brings extreme standardization. It comes back to the need to maintain complexity when dealing with human questions."
- A participant

> A social justice and equity principle would involve making AI benefits available to all, and that AI development will not contribute ton increasing social and economic inequalities, but rather reduce them.

"Accessing new technologies: a privilege only the rich can afford? Will this type of technology widen inequalities?" - A participant

ISSUES OF TRUST, RELIABILITY, SECURITY: A CAUTION PRINCIPLE?

Furthermore, the issues concerning trust in the development of AI technologies were regularly brought up. The issue of trust in AI and its algorithms in different sectors mainly presents itself as a certain suspicion towards these techniques as well as how representative the selected data and the validity of the interpretations made really are, suggesting a caution principle:

"Since it's scientific, a person could tend to forget that an algorithm can be wrong: caution." - A participant

This trust issue is also closely tied to the question of the reliability of AI systems. Paying close attention is paramount to ensuring the quality of the collected data and the correlations that can be made as well as their purpose in order to avoid blind faith and prevent potential manipulation.

Along the same lines, the participants raised issues tied to the security of AI devices, namely the risk of potential abuse, pirating and cyberattacks on the systems and the data they hold, as well as the validity of the recommendations and decisions made by algorithms. These issues are accompanied by a dilemma between "collective fluidity" and "system vulnerability", meaning AI that must be both flexible and solid (ex. in the smart city and connected objects sector).

TOWARDS A TRANSPARENCY PRINCIPLE?

At the crossroads of the knowledge, responsibility and justice principles lies a transparency principle that implies being able to understand an algorithmic decision and react to it. For that to happen, citizens must insist on algorithmic procedures being explainable so that anyone can understand and verify the criteria that was taken into consideration when making the decision:

"Transparency in the variables used, the data, the parameters. Explaining a decision in clear, natural language."

(Workplace Table, Bibliothèque Mordecai-Richler, Montreal, March 10, 2018, AI as mandatory path to the workforce scenario).

This explainability issue implies the necessity of finding a way to simplify these algorithmic procedures so anyone can make sense of them; this goes hand-in-hand with the development of digital literacy, which will enable enlightened consent and a critical mind towards the system. The explainability of these algorithms accompanies the issue of being able to verify algorithmic decisions, hold someone responsible for those decisions and eventually correcting certain negative effects such as discriminatory biases. It's also about making these algorithm explanations accessible, for example through open development (open source, licence free, open data), namely out of a concern for feedback to understand why a decision was made and manage eventual feelings of injustice after being refused (when applying for a job, social assistance or insurance, for example).

A warning was made on two occasions, however, regarding this transparency principle: This transparency could have a potential effect on the security of the algorithmic systems (risk of hacking). This transparency principle would tie into the issue of trust in AI technologies.

"If AI analyzes things that are too complex for the human brain, who's keeping an eye on what's going on behind the curtain?"
- A participant.

6.4.3

ISSUES REGARDING THE RELATIONSHIP BETWEEN HUMANS AND AI

Regardless of the sector, the citizens identified many issues regarding the relationship between humans and AI. Participants namely worried about the place left for humans in such systems, which could lead to various abuses.

For instance, citizens are concerned with the respect of "human nature", across all sectors. It is confronted with the place of the object in society and its relationship to the human for the smart city and connected objects sector: will reduce the status of the human being and grant more importance to protecting objects? Respecting "human nature" also means, in the eyes of the citizens, guaranteeing to take into consideration some singularity, some complexity, some human messiness according to numerous parameters that are hard to quantify, such as what the participants called "individual charisma" in the legal sector. Not taking into account human dynamics and its possibilities for change shows a concern with the "static" vision of a human being provided by the algorithm, which would make its decisions problematic and unreliable. In healthcare, this reliability is also questioned when it comes to a diagnosis or suggestions provided by algorithms that have no holistic visions of the individuals, who can't be reduced to their biological data.

Participants are worried about an eventual dehumanizing of services which could appear if AI is granted too much space. In healthcare, it's a certain dehumanizing of care and the loss of the doctor-patient relationship. In the smart city and connected objects sector, participants are concerned about striking a balance for a harmonious development of society and human beings while implementing AI and connected objects. In the workplace sector, this dehumanizing can be perceived as the automation of tasks. In the legal sector, it could come from a potential lack of "empathy", "instinct", "wisdom" in AI systems, which raises concerns about prosecuting cases rather than treating them "humanly":

> *"Cases will become standardized and the person themselves won't be considered enough."*
> - A participant

Participants are worried about a loss of emotional and relational quality, sometimes seen as a potential "denaturalizing", even an "alienation" (from social life in favour of digital life), this across all fields. These concerns namely refer to the transformation in the relationship to care, knowledge, wisdom, work, but also the skills of individuals.

> *"The challenge isn't making machines more intelligent, it's making humans more intelligent."* - A participant

Will doctors still have the same expertise if they're constantly relying on expert systems? What effect will that have on the trust placed in their expertise as opposed to the AI's? A similar reflection was brought up in the education sector regarding replacing teachers with AI:

> *"If there's too much AI equipment in schools, teachers will become useless."*
> - A participant

Still in the education sector, the citizens reminded everyone that human intervention is necessary:

> *"We can't rely solely on a machine."*
> - A participant

Finding a path to complementarity between humans and AI therefore seems very important. This complementarity was discussed as a "good balance" for sharing tasks (for example, between the "objective" and "subjective" for the workplace sector, between an infinitely patient AI as a "learning assistant" and a teacher with emotional and relational capacities for the education sector). In the workplace sector, citizens suggested implementing a watch to preserve "human primacy", which should guarantee that technology is only a support:

> *"A guarantee that the system is not an end in itself, but that it is focused around the human."*

(Workplace Table, Musée de la civilisation, Quebec City, April 6, 2018, AI as mandatory path to the workforce scenario).

6.5

RESULTS: POTENTIAL SOLUTIONS

6.5.1

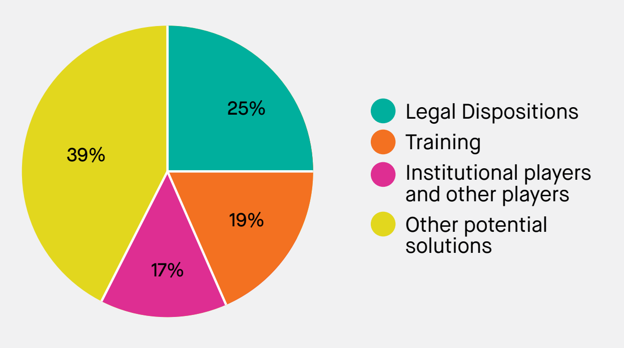
THE GENERAL POTENTIAL SOLUTIONS SUGGESTED BY CITIZENS

Citizens who took part in the co-construction days were invited to suggest potential solutions to the previously identified issues. 190 potential solutions were formulated and adopted through consensus during these activities (although other suggestions may have been discussed during the tables.)

All co-construction tables agreed on 3 general potential solutions to guarantee socially responsible AI development, regardless of sector:

1. Legal dispositions
2. Putting training in place for all
3. Identifying independent key players for AI management.

Figure 8: Three general potential solutions at all tables



Regardless of the sector, all tables agreed on recommending implementing a legal framework adapted to the reality of AI development and personal data management (especially massive data). For example, participants recommended implementing specific rules and laws, new types of contracts, even putting a moratorium in place. Implementing training that is accessible to all was also strongly recommended, both for professionals of the affected sectors (to guarantee adequate use of AI systems in their work) and the general population (to guarantee everyone can participate in the debate and gain basic digital literacy).

Citizens also identified the institutional players and the key independent and competent players (existing or to be created) who would oversee the responsible development of AI. The players identified are people (ex. ombudsman, auditor, life and well-being commissioner) or groups of people (ex. setting up an artificial intelligence centre for civilian security, a 1–800 number against connected objects discrimination or a Ministry odf data ethics and digital protection).

In all sectors as well, citizens suggested creating technical and ethical evaluation mechanisms for AI. Namely, establishing a certification (or label) system as an ethical guarantee was suggested on many occasions. Different tables also recommended implementing a code of ethics (whether it's a matter of updating the existing code or creating new ones); and participatory mechanisms (ex. co-constructions, public consultations or an AI summit) in order to guarantee a democratic development of AI and its management. The importance of implementing research programs in various disciplines (ex. philosophy, social sciences, bioethics) was also raised. The creation of digital tools (ex. digital and interactive healthcare forms, individual digital file in the workplace sector) was also suggested.

Developing incentives that aim to encourage responsible development – was agreed upon at different tables, as was implementing diversity quotas (which reward companies that guarantee not to exclude or discriminate against certain minorities through AI biases) or funding companies that establish transitions for employees whose job is being replaced by AI. Finally, establishing professional frameworks (and different internal procedures for companies) and the creation of public policies that could lead, for example, to the creation of a digital citizenship, were all put forward.

Table 8: Number of tables suggesting each category of potential solutions

	Education	Legal system and predictive police	Healthcare	Workplace	Smart city and connected objects	Total
Legal dispositions	6	7	8	3	10	34
Training	4	5	6	4	10	29
Institutional players and other players	1	6	7	3	9	26
AI evaluation devices	1	3	8	1	5	18
Code of ethics/ conduite	5	2	4	2	2	15
Participative mechanisms	2	2	1	3	5	13
Research programs	1	2	4	1	1	9
Digital tools	0	0	1	2	3	6
Professional frameworks and internal policies	1	1	2	0	0	4
Incentives	0	0	0	2	1	3
Public policies and guidelines	1	0	0	1	1	3
Number of co-construction tables	9	8	12	5	11	45

6.5.2

POTENTIAL SOLUTION BY SECTOR

EDUCATION

Citizens gathered around 9 co-construction tables in which the theme of AI development in the education sector was discussed. Participants formulated 27 potential solutions or general AI framework guidelines during these activities.

Table 9: Potential solutions or general guidelines for the education sector

	Number of potential solutions formulated
Legal dispositions	8
Training	7
Code of ethics/conduct	5
Participative mechanisms	2
Institutional players and other players	1
AI evaluation devices	1
Research programs	1
Professional frameworks and internal policies	1
Public policies and guidelines	1
Total	27

7 POTENTIAL SOLUTIONS CONCERNING TRAINING WERE FORMULATED BY 4 TABLES OUT OF 9:

> TRAINING

In regards to education, participants recognized the need to be proactive in setting up training for the entire community affected by AI development in that sector. This training should cover digital literacy, media literacy, as well as ethics and the issues tied to integrating AI in an educational environment. This training could, for example, take the form of digital literacy accompaniment for both parents and students, or be directly integrated into the initial citizen training.

The citizens also recommenced training education professionals more specifically, for instance by including the development of work skills "teamed up" with AI devices in the curriculum for the initial and university training of teachers (ex. a certification for the B.Sc. or an accreditation system). This training will have to be both technological (how to use AI), but also geared towards teaching techniques with AI (how to organize teaching sequences and insisting on the fact that knowledgeable professionals orchestrate AI, not the other way around.

"Accrediting agents of change (both psychoeducators and active teachers) by teaching establishment to gradually integrate AI in an academic environment."

(SAT Table, Montreal, March 13, 2018, AlterEgo scenario).

The importance of establishing adequate training was also raised. The training's purpose would be to provide the appropriate information allowing stakeholders to accept their responsibility towards AI, in order to avoid teachers putting blind faith in educational AI devices. This training would accelerate the understanding of actors in the field of education and favour their mobilization to develop AI so it serves the autonomy of the learners while preparing them to deal with these realities. This training will help develop human skill sets and provide power to guide and even redefine future AI development.

"Raise awareness around responsible use of AI and promote a diversity of relationships to knowledge."

(SAT Table, Montreal, March 13, 2018, Nao scenario).

LEGAL SYSTEM AND PREDICTIVE POLICE

Citizens gathered around 8 co-construction tables to discuss the theme of AI development in the legal sector. Participants formulated 36 potential solutions or general AI framework guidelines during these activities.

Table 10: Potential solutions or general guidelines for the legal system and predictive police sector

	Number of potential solutions formulated
Legal dispositions	**10**
Institutional players and other players	7
AI evaluation devices	5
Training	5
Code of ethics/conduct	2
Participative mechanisms	2
Research programs	2
Professional frameworks and internal policies	1
Total	34

10 OF THE POTENTIAL SOLUTIONS FORMULATED ARE LEGAL DISPOSITIONS AND ARE RECOMMENDED BY 7 TABLES OUT OF 8:

> LEGAL DISPOSITIONS

In regards to the legal system and predictive police, it is imperative to establish laws and regulations on transparency: it's a matter of demanding transparency from private and public companies collecting criminal data, but also of laying bare the decision-making processes when these decisions are made by algorithms. Explaining the decision must come with measures allowing access to mobilized algorithms and ensuring they are explained in intelligible fashion. As a first transparency mechanism, many participant tables suggested that the AI used in the legal sector—even all public sector AI—be developed in open code, under free licence. From a legal standpoint, it's about guaranteeing "the right to a full answer and defence", namely with the possibility to challenge a decision by raising procedural or formal deficiencies (Table Musée de la civilisation Table, Quebec, April 6, 2018, Parole scenario).

This transparency imperative goes hand-in-hand with establishing legal dispositions giving the right, believed to be fundamental, to be judged by a human being to preserve procedural justice and individualization of the sentence. Underlining the need for law to adapt to a new technological reality with AI in legal decision-making, many debates occurred around conciliating human and artificial players in this process. The consensus was as follows:

"The right to appeal before a human judge: The appeal procedure for a decision made by a computer must always be heard by a human judge."

(Musée de la civilisation Table, Quebec City, April 6, 2018, Parole scenario).

In the perspective of preventive AI used for police purposes, it is mentioned that there is a desire to establish a "framework that allows us to go beyond and eliminate biases, discrimination and abuse of power" (SAT Table, Montreal, March 13, 2018, Predictive Arrest scenario) as well as reinforce laws around consent to ensure it is truly an enlightened one. There's also the idea of limiting public and private stakeholders access to private data such as "private conversations on digital platforms" (Du Boisé Library Table, March 17, 2018, Preventive Arrest scenario) and enforcing a "right to be forgotten, to modify and correct data as well as a right to personal access to the data gathered" (Père Ambroise Library Table, March 3, 2018, Predictive Arrest scenario).

HEALTHCARE

Citizens gathered around 12 co-construction tables to discuss the theme of AI development in the healthcare sector. Participants formulated 46 potential solutions or general AI framework guidelines during these activities.

Table 11: Potential solutions or general guidelines for the healthcare sector

	Number of potential solutions formulated
Legal dispositions	11
Institutional players and other players	9
AI evaluation devices	**8**
Training	6
Code of ethics/conduct	4
Research programs	4
Professional frameworks and internal policies	2
Participative mechanisms	1
Digital tools	1
Total	46

8 FORMULATED POTENTIAL SOLUTIONS ARE AI EVALUATION DEVICES (IN HEALTHCARE, CERTIFICATIONS), AND ARE RECOMMENDED BY 8 TABLES OUT OF 12:

> AI EVALUATION DEVICES

Citizens recommended establishing AI ethical certification in healthcare, meaning the development of a certification (or label) for algorithms and robots, on the database from research projects (participative study on the context that influences AI development) to determine the criteria for this certification and its various levels. These criteria should include transparency, security and relevance of the tool. For example, these certifications would be designed to standardize access to the decision-making process of the algorithms, or to validate the tools of healthcare robots. These certifications should be issued by the government or independent, multiparty organizations to protect public interest and patient well-being, would mainly target private companies developing AI healthcare.

“Upfront certification for healthcare robots and their toolbox (namely, to protect public interests)”

(Mordecai-Richler Library Table, Montreal, March 10, 2018, Helper robots for the elderly scenario).

WORKPLACE

Citizens gathered around 5 tables to formulate 32 potential solutions regarding AI development in the workplace.

Table 12: Potential solutions or general guidelines for the workplace sector

	Number of potential solutions formulated
Training	8
Institutional players and other players	5
Legal dispositions	5
Incentives	3
Participative mechanisms	**3**
Code of ethics/conduct	2
Digital tools	2
Public policies and guidelines	2
AI evaluation devices	1
Research programs	1
Total	32

THE SUGGESTIONS CONCERNING PARTICULARLY STOOD OUT. THEY WERE RECOMMENDED BY 3 OUT OF 5 TABLES:

> PARTICIPATIVE MECHANISMS

Participants suggested creating a multi-sectorial "permanent consultation space" within the government, to respond to the division of powers (tied to the democracy principle). The information gathered digitally could then be more accessible and that space would be responsible for structuring sectors emerging in the field of employment.

Citizens also mentioned the importance of user participation in designing the interface of AI tools, which could take the form of "design thinking" with different partners and would allow them to review the work of the programmers:

"Allowing user input in machine learning through open AI (based on the Wikipedia model) to correct and review biases by and for society."

(Musée de la civilisation Table, Quebec City, April 6, 2018, AI as mandatory pathway to employment scenario).

User feedback should help follow data collection and algorithm development, and reduce the "gaps" that could lead to prejudice towards individuals from competent authorities (ex. ethics committees, corporations) to adapt the system.

SMART CITY AND CONNECTED OBJETS

Citizens gathered around 11 tables on the theme of AI development in the smart city and connected objects sectors. These 11 tables formulated 51 potential solutions.

Table 13: Potential solutions or general guidelines for the smart city and connected objects sector

	Number of potential solutions formulated
Legal dispositions	13
Institutional players and other players	**10**
Training	10
AI evaluation devices	5
Participative mechanisms	5
Digital tools	3
Code of ethics/conduct	2
Incentives	1
Public policies and guidelines	1
Research programs	1
Total	51

> INSTITUTIONAL PLAYERS AND OTHER PLAYERS

Table participants discussing the theme of smart city and connected objects suggested many ideas for the creation of institutional players, whether independent societies or advisory committees. The democratic ideal of committees or assemblies allowing citizen participation was recalled many times.

For the control of connected objects, 2 models were therefore suggested, including a mechanism forcing the self-regulation of private players:

> Based on the model of the Régie du logement du Québec, a Régie des objets connectés (connected object management) would help set prices for connected objects (such as refrigerators) and would set forward social assistance to facilitate their acquisition.
It would also issue ownership certificates when purchasing a connected object to establish that the data generated by this object belongs to the user. This person can then choose to give their consent or not for the data to be communicated to the company commercializing the object as well as their insurance, without risking any penalties.

> An independent authority on data management could allow citizens to conduct a class action when there are abusive uses. It could also manage a digital platform where users can speak freely and publicly about the advantages and disadvantages of AI devices and thereby have an impact on the branding of private players commercializing these devices. The private players would then be forced to self-regulate through the pressure users place on their image (Musée de la civilisation Table, Quebec City, April 6, 2018, Connected refrigerator scenario).

To respond to an equity issue and thereby ensure an equitable sharing of AI, an advocate could be reached at "1–800 discrimination of connected objects" (INM Table, Montreal, February 18, 2018, Connected refrigerator scenario). It could then be a part of a "multiparty committee that democratically manages incidents, injustices and other issues" (Mordecai-Richler Library Table, Montreal, March 10, 2018, Self-driving car scenario). Furthermore, an independent auditor could be mandated to lead an accounting audit to ensure an equitable sharing of AI benefits (INM Table, Montreal, February 18, 2018, Connected refrigerator scenario).

For self-driving car regulation, the creation of the SAIAQ (Société de l'Assurance de l'Intelligence Artificielle du Québec) would bring modifications to road safety laws to adapt them to autonomous driving. It would also include auto insurance 2.0 that would suggest new kinds of contracts for this type of driving (Bibliothèque du Boisé Table, Montreal, March 17, Self-driving car scenario).

7. A CONTINUOUS CO-CONSTRUCTION PROCESS

7.1

CONTINUE THE DELIBERATION

The *Montreal Declaration* project concentrated its first phase on five key sectors: education, health, work, smart city and predictive police. An entire year of co-construction wouldn't even cover all the reflection themes. The co-construction initiative will therefore continue in September 2018, allowing for discussions about new themes that had barely been touched upon in the scenarios used in the co-construction phase.

Among these:

ENVIRONMENT

Environmental issues, as we all know, are vital issues that humanity must face in the near and far future. AI can help optimize the use of our resources, but it also uses a great deal of resources and energy, and generates electronic waste. AI development is also an environmental issue, but the future of protecting the environment will also come through a targeted use of AI. What choices must we make as a society to fulfill our environmental obligations?

DEMOCRACY AND MEDIA PROPAGANDA

AI and the use of megadata present major challenges for democracy as a political participation system. As soon as you think of social media, for example, the scandal of using user data for political purposes and the use of chatbots to massively spread fake news and contaminate the electoral process immediately comes to mind. Our relationship with the integrity of information sources and media is deeply affected. How to guard against an adverse use of AI in the political arena? How do you guarantee the conditions for every citizen to express their critical autonomy?

SECURITY AND INTEGRITY

To discuss the issues concerning the development of autonomous weapons, of intervention police inside a country (where the recommendations have the best chance of making an impact), double use and misuse of AI, data integrity (ex. protection against cyberattacks, etc.).

ARTS AND CULTURE

In this era of digital technology, we elaborate, discover, explore new ways to produce cultural and artistic objects. The impact of AI can be felt on both artistic creation and circulation of the work.
An AI culture is also developing, and our relationship to other cultures is modified because of it.

PREDICTIVE JUSTICE

Although the question of justice was discussed in the conditional parole scenario, it's worthwhile to discuss it further by questioning the future of law and legal rulings, when algorithms are used to predict a legal ruling. This field of algorithmic development is starting to disrupt the practices of judges, lawyers and mediators, but it could also modify the way law is shaped, especially jurisprudence.

Co-construction workshops on these themes will produce a series of analyses and suggestions that will complete those that were produced in the first co-construction phase. They will also feed the reflection on the ethical principles of the *Montreal Declaration* and the recommendations for a public policy on AI which will be developed in their extension.

We will present public policy recommendations around priority fields of action. The priority fields of action are transversal recommendation axis with sectors and themes. We will only reveal the priority fields, and the recommendations, once the deliberation process is complete, but we can already say that three fields of action have established themselves:

> Digital literacy

> Diversity and inclusion

> Transition and social mutations

7.2

AN INSPIRING INITIATIVE

Among the initiatives inspired by the *Montreal Declaration*, we must first mention the work of the Montreal AI Ethics Meetup group, founded and coordinated by Abhishek Gupta (McGill). This group, which brings together over one hundred multidisciplinary researchers and concerned citizens concerned by AI developments, devoted many 2h sessions, between December 2017 and March 2018, to the Declaration principles. Although they are not citizen deliberations, the Montreal AI Ethics Meetup sessions are nonetheless true collective intelligence exercises that involve a variety of high-level researchers. A detailed report of their critical reflections was submitted to the Declaration team and can be read online. The authors of the report are: Stephanie Dyke, Paule-J Toussaint, Abhishek Gupta, Gregory Caicos, Marc Daher, Peter Chen. This initiative is especially encouraging because it comes from the heart of Montreal's AI community.

Another important initiative was the evening of reflection organized by ESG UQÀM (École des Sciences de la Gestion de l'UQAM) on February 15, 2018, entitled: "Vers un développement responsable de l'IA : Soirée de réflexion autour de la Déclaration de Montréal pour un développement responsable de l'ia" (Towards a responsible development of AI: Evening of reflection around the *Montreal Declaration*). This evening brought together seven researchers from UQÀM around the seven values of the Declaration. After a general introduction by Yoshua Bengio and Martin Gibert, the audience had a chance to hear the thoughts of professors Marie-Jean Meurs (well-being), Christophe Malaterre (autonomy), Hugo Cyr (justice), Sébastien Gambs (privacy), Étienne Harnad (knowledge), Dominic Martin (democracy) and Maude Bonenfant (responsibility). The summary of the exchanges that took place during this university meeting can be read online[21].

At Université de Montréal, the Faculty of Arts and Sciences created Perspective in February 2018, an interdisciplinary lab of ideas whose first is explicitly aligned "in the tracks of the *Montreal Declaration*". This lab brings together a group of graduate students tasked with producing reports with the intention of enlightening public policymakers of the social impacts of AI.

Finally, many organizations (businesses, development organizations or associations) showed their interest in the *Montreal Declaration*, hosted presentations of the Declaration or organized discussions about its principles. This is the case for IBM, Montréal InVivo, l'ACFAS, Printemps numérique, or C2Montréal.

The citizen involvement and consultation process initiated by the *Montreal Declaration* is growing and is now operating outside Quebec, in Toronto under the impulse of the ICRA, and also soon in Europe, Brussels and London.

[22] https://docs.wixstatic.com/ugd/ebc3a3_0e7d08f785c54b148d34c1c6c54f4b8c.pdf

CONCLUSION

The new possibilities introduced by AI should deeply transform society in the coming years. But how to ensure that these algorithms and data sets lead to positive social innovations in health, education, the justice system, public and private organizations, in our cities and our everyday lives? Between the many promises of AI and the very real ethical risks, how do you see straight? And how do you debate it?

The co-construction approach of the *Montreal Declaration* for responsible AI was first born from an observation: the debate around responsible AI concerns us all and the feeling of uncertainty and concern provoked by AI is a shared one. This debate therefore cannot be left only to the experts. From this observation, the co-construction approach took a chance that a debate involving citizens, experts and stakeholders was not only possible, but could potentially generate very innovative ideas, as long as it is well organized.

To that effect, this intermediary report drafted after more than ten workshops from February to May 2018 in Montreal, Quebec and Laval, presents an overview of two expected results at the end of the co-construction process in December 2018: first on the processes implemented to organize the co-construction, then one on the first thematic recommendations formulated by the participants in these meetings.

On the participation processes used to organize this debate, the 3h world café formula in public libraries, open to all, like the one on the big day with stakeholders, citizens and experts (in Montreal and Quebec), both generated very rich exchanges. In particular, the choice of inspiring public spaces to hold these free events (public librairies, Musée de la civilisation in Quebec, Société des arts technologiques in Montreal), the attention devoted to create good humour among the participants, to providing them information, knowledge at the right times (Contextual introduction at the start of the workshop and handout of a participant's guide presenting the *Montreal Declaration* and deliberation resources during the workshop), the use of prospective scenarios presenting user AI scenarios set in 2025 in Quebec, providing participants with "anchors" and "triggers" for the debate while suggesting frameworks likely to avoid certain cognitive biases, facilitating discussion through reflexive facilitators, sharing the way of facilitating this type of prospective deliberation throughout the event, using signs to accentuate and synthesize the intermediary results of the deliberations, all of this helped implement a workshop that was friendly, welcoming and a source of many relevant and innovative recommendations in a limited timespan.

This type of workshop will continue over the coming months, the idea being to experiment with various devices and supports for the debates. In particular, the project of developing a serious board game to stimulate ethical and prospective deliberation around AI to explore and test future workshops.

After this first step of deliberation with citizens, researchers and stakeholders, many issues and potential solutions were formulated by hundreds of citizens gathered around co-construction tables. As the objective of the initiative was to stimulate citizen deliberations on the responsible development of AI, discussions were organized around scenarios presenting fictional decision-making situations, ethical issues, possible risks or controversies, to both exemplify and test the principles of the *Montreal Declaration* on responsible AI.

These first results give a certain idea of the social acceptability of both AI and its development. For the purposes of this preliminary analysis, we've chosen to stay as close to the voice of the citizens and stakeholders who took part in the debates as possible. Furthermore, the results presented at this halfway point of the co-construction process (issues, potential solutions, potential new principles) are not definitive and will be explored in greater detail in the next steps of the analyses, namely on the conditions of the implementation.

This first phase of consultations also highlighted certain dilemmas to explore. For example, in the field of healthcare, data confidentiality protections butt heads with the promise of predictive, preventive and personalized healthcare, which would require AI to take a great deal of data into consideration (and not only biological data). Another example, in the fields of health, education, or the legal system, although there is a consensus to say that humans must remain masters of the decision, in practice this can run up against a manager's will to automate certain decisions in order to increase productivity in his organization. As for measures to adopt, opting for legal measures to provide a framework for businesses can be confronted with the need to support rapidly growing, innovative companies that can contribute to a country's prosperity in the future. These dilemmas are challenges for the responsible management of AI in society, and each can be a starting point for a collective innovation process to imagine unprecedented ways of overcoming them.

The development of AI raises many ethical and societal questions that the co-construction process will continue to analyze over the coming months, by expanding on the sectorial issues explored in this first step, and discussing new themes: the environment and energy transition, the relationship between democracy, propaganda and media, security, autonomous weapons and data integrity, arts and culture. All of these transformations brought on by the development of AI in different social spheres make us question, as citizens, what kind of society we should build. At the heart of the tensions between hopes and fears, it's the interactions between humans and technology that it will be essential to watch and analyze in prospective and critical fashion. If one recommendation was unanimous in the co-construction debates, it is indeed keeping a central role for humans in a world that grows more and more artificially intelligent.

PARTICIPANTS IN THE CO-CONSTRUCTION

Thank you to the students and professionals that facilitated the workshops and took notes:

Alexandre Beaudoin-Peña, Université de Montréal
Bhavish Beejan, Université Laval
Karl Bherer, Université Laval
Alexis Bibeau, Université Laval
Pierre-Antoine Boutin-Panneton,Université Laval
Dominic Cliche, Université Laval
Eve Gaumond, Université Laval
Emilie Guiraud, Université Laval
Hubert Hamel-Lapointe, Université de Montréal
Audrey Houle, Université Laval
Nico Julien, Université Laval
Henri Lajeunesse, Université Laval
Guillaume Macaux, Université Laval
Mariève Mauger-Lavigne, Université de Montréal
Orly Nahmias, citizen
Judith Paquet, Université Laval
Pierre-Luc Plante, Université Laval
Lynda Robitaille, Director of Operations at Centre de recherche en données massives (CRDM)
Jason Stanley, Université de Montréal
Yanis Taleb, Université de Montréal
Clémence Varin, Université Laval

Thank you to the citizens, the professionals and the experts that took part in the workshops (only the names of the participants who gave us their consent are published below):

Sihem Neila Abtroon
Béatrice Alain
Hassane Alami
Rana Alvabi
Alejandro Arreola-Alvarado
Gabriel Arruda
Barthélémy Aucourt
Naomi Ayotte
Manon Babine
Maryluisa Barillas
Philippe Beauchemin
Stéphane Beaulieu
François Beauregard
Claude Bédard
Sylvain Bédard
Vincent Bergeron
Alexandre Berkesse
Karl Bherer
Marise Bonenfant
Serge Bouchard
Caroline Boudreault
Lyne Bourbonnais
Véronique Boutier
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
Lilen 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 Décary-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
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
Christian Goudreau
Gilles Gouin
Mervine Gowry
Alexandre Gravel
Michel Grou
Alexandre Guédon
Pascaline Guenou
François Guité
Carl Hamilton
Simon-Pierre Harvey
Lucie Hébert
Ghiles Helli
Lucas Hubert
Aida Issa
Sabrina Jocelyn
Erwan Jonchères
Nico Julien
Debbie Jussome
Ed Khazen
Amy Khoury
Andrée Labrie
Anne-Marie Lacombe
Marie-Claude Lagacé
Henri Lajeunesse
Karine Landry
Jean-Michel Lapointe
Jonathan Lasprilla
Sylvie Lavoie
Louis Lecaer
Dominique Leclerc
Sarah Legendre Bilodeau
Pascale Lehoux
Claude Lejeune
Mélanie Levasseur
Elisabeth Limoges
Robert Locas
Santiago Lopez
Aurélie Macé
Aicha Mafhoum
Suzanne Mainville
Mantas Manovas
Mathieu Marcotte
Jean-Pierre Marquis
Marie Martel
Mariève Mauger-Lavigne
Moussa Mekhnach
Natacha Mercure
Bruno Milia
Michael David Miller
Ann Mitchell
Erica Monteferrante
Farida Mostefaoui
Maria Moudfir
Vanessa Murray
Orly Nahmias
Vanessa Nantel
John Newhouse
Justin Ngoza
Zoonie Nguyen
Catherine Olivier
Daniel Pascot
Florence Paulhiac
Jorge Perez
Lorenzo Perozzi
Geneviève Perreault
Benoit Petit
Louis Piette
Frédérick Plamondon
Pier-Luc Plante
Kamila Podgorska-Gilbert
Keith Poitras
Julie Politi
Thomas Poulin
Denis Poussart
Emmanuelle Praine
Louis-Philippe Pratte
Mariel Ramos
Diane Raymond
Catherine Régis
Laurence Renault
Cassie Rhéaume
Toussaint Riendeau
Anne-Marie Robert
François-Xavier Robert
Louis-Nicolas Robert
Nicolas Roby
Stéphane Roche
Marie Roy
Sara Russo-Garrido
Laurence Sabourin
Iger Sadoune
Marie-Noëlle Saint-Pierre
James Sangster
Sylvie Saucier
Sébastien Adam
Jean-François Sénéchal
Eric Shannon
Danielle Sicotte
Chantale Simard
Julie Simard
Jean-Hébert Smith-Lacroix
Karima Smouk
Yanis Taleb
Isabelle Tanba
Christian Tanguay
Marc Tomkinson
Daniel Tremblay
Jérémy Trudel
Marie-Christiane Trudel
Félix Vaillancourt
Julie Verdy
Danael Villeneuve
Grant Wark
Bryn Williams-Jones
Lemy Wong
William Wong
Almina Yagoubi
Ming Yue

ANNEXES

ANNEX 1 *Co-construction Workshops: Detailed description and Procedure*

WORLD CAFÉS

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

The world café is an enjoyable conversation device that seeks to facilitate constructive dialogue and the exchange of ideas. We seek to recreate the ambiance of a café 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 from the new arrivals. The ongoing conversations are therefore "pollinated" by the ideas of the previous conversations. At the end of the process, the main ideas are summed up during a plenary assembly, and possible follow-ups are submitted for discussion[22].

This world café technique was adapted and enriched with many elements:

- An introduction to the *Montreal Declaration* and the social and ethical issues of AI;
- the reading of prospective sectoral scenarios set in 2025 to spark the discussion;
- the use of a poster to document the discussions;
- the handout of a participant workbook presenting the principles of the *Montreal Declaration for Resonsible AI*, a lexicon and an exemplified typology of possible recommendations.

Here is what a typical world café looks like:

Table 14: Typical procedure for world cafés

Steps	Time	Description
Welcome	1 pm to 1:30 pm	Coffee and snacks
Discovering AI and its ethical and social implications	1:30 pm to 2 pm	**Educational Introduction:** introduction to the ethical and social implications of artificial intelligence (*Montreal Declaration*), presentation of scenarios set in 2025 and of the activity.
World café	2 pm to 4 pm	- 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 a responsible rollout of AI.
Summary in plenary session	4 pm to 4:30 pm	Summary of the discussions in plenary session The facilitators sum up the posters from each thematic island, followed by a group discussion.

[22] Definition from the Institut du nouveau monde (INM)

CO-CONSTRUCTION DAYS

These one-day meetings brought together citizens, stakeholders and experts that seek to further explore sectoral issues and develop recommendations. They rely on the prospective co-design model, developed at the University of Montreal's Lab Ville Prospective.

The prospective co-design model relies on many principles, at the crossroads of design, participation and forecasting: the mobilization of typical scenarios and unknown prototypes as conversation starters, means of abandoning cognitive fixation, and exploration vehicles (that's the design dimension); Collective participation devices bringing together players from multiple horizons, citizens and organizations as experts (for the collective aspect of the "co"); lastly, the forecasting approach which consists of projecting oneself into a possible future 10 or 20 years down the line to perform 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 the otherness of forecasting: according to him, "the future engages the present on the alterity mode". And Georges Amar, in an article on conceptive forecasting (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 function of forecasting is to work on the unknown, to put 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 day looks like:

Table 15: Typical procedure for co-construction days

Steps	Time	Description
Welcome	8:30 to 9 am	Coffee and pastries
Introduction and AI Discovery	9 am to 10 am	Introductions: principles of artificial intelligence, ethical issues surrounding AI (*Montreal Declaration*) and forecast scenarios.
Team forecast	10 am to 11:30 am	Team forecast: starting with a trigger scenario and the *Montreal Declaration* principles, formulate the ethical and social issues raised by the 2025 scenario and explore how an ethical controversy could appear or grow.
	11:30 am to 12:30 pm	Plenary: Plenary presentation of ethical and social issues raised for 2025, and discussions with the group as a whole.
Lunch on site	12:30 pm to 1 pm	Lunch
Developing recommendations	1:30 pm to 2:45 pm	Developing recommendations. Work in teams: using the 2025 ethical issues identified in the morning, develop recommendations (rules, sectoral codes, labels, public policies, research programs, etc.) to establish starting in 2018–2020 in Quebec.
	3 pm to 4 pm	Plenary team presentations and group discussion
Conclusion and follow-up	4 pm to 4:30 pm	Review and observations surrounding the day

ANNEX 2

Prospective scenarios

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 the debates and deliberations on the ethical questions raised by artificial intelligence. The 2025 horizon was selected to be in the near future, at the heart of the 2020–2030 decade which should be the one that sees an intensive rollout of artificial intelligence in society.

1. EVERY SCENARIO SUMMED UP BY THEME

From February to May 2018, eighteen scenarios were debated. The table below presents a brief summary of these scenarios.

Table 16: Scenario summaries

Theme	2025 AI scenario	Summary of AI scenario in 2025 in Quebec
1. Predictive Health	Healthy digital twins	Olivier learns that one of his 126 digital twins has received a depression diagnosis. Should he go see a professional?
	Discriminating Health Insurance	Olivier's insurance company asks him to change his lifestyle, based on his personal data. Can he refuse without any consequences?
	Vigilo, a House Robot for the Elderly	Soline is 80 years old and lives at home with Vigilo, his robot companion. This one regularly reports predictive diagnoses on Soline's health to her family. Does she wish to have everything revealed?
	A therapeutic decision at the hospital	An experienced doctor and a medical recognition algorithm don't quite agree on a diagnosis.

Theme	2025 AI scenario	Summary of AI scenario in 2025 in Quebec
2. Smart City	Self-driving cars (setting the algorithm and sharing the road)	To guarantee 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.
	Self-driving cars (restricted use)	Self-driving cars have become a ride-share service for citizens. Priority access criteria is managed by AI in order to maximize the city's predictive economic growth.
	A connected fridge that wants what's best for you (nudges)	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?
	A social rating based on a carbon footprint	A family's consumption is defined and tracked in order to prevent a negative impact on the environment.
	A smart toy that's not all that loyal!	How far does a smart toy's loyalty to a child go? Is it the same as a friend's?
3. Predictive education	AlterEgo, AI that assists learning at school	AI helps students learn more efficiently, thanks to personalized homework and exercises. Does the teacher still have complete professional autonomy?
	AlterEgo2, AI School Guidance Assistant	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?
	Nao, AI that helps prepare conferences	AI helps a lecturer develop his presentation and update it throughout the lecture, according to the reactions of his students.
4. Police and predictive justice	A preventive arrest in a public space	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.
	A Parole Decision	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).

Theme	2025 AI scenario	Summary of AI scenario in 2025 in Quebec
5. Workplace	AI to optimize workplace atmosphere	A company's human resources department uses AI with data mining to evaluate the behavioural style of their employees and guide them towards the standard "good workplace atmosphere".
	Recruitment AI as a mandatory path to the job	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?
	Socially Responsible Structuration	A sustainable logistics company must massively incorporate AI into many of its services in order to remain competitive. But it wishes to do so in socially responsible fashion.
	A New Committee on Professional Development	A company's professional development committee welcomes new members: the representatives of the collaborating robots. Not everyone shares the same opinion on this evolution.

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 *Montreal Declaration*: predictive health, predictive education, smart city, predictive justice, and the transversal theme of transformations in the workplace.

Each scenario presents the story of a case that was built by combining many dimensions: a sectorial problem, a user experience set in 2025, a learning apparatus mobilizing data and one more artificial intelligence techniques, and finally, ethical and social issues.

Table 17: Elements of five scenarios

2025 AI Scenarios	Digital twins	Self-driving cars	AlterEgo	Parole	Responsible Restructuration
Themes	1. Predictive healthcare	2. Smart City	3. Predictive education	4. Predictive justice	5. Workplace
Sectorial problem	Preventive healthcare and personalized by similar profile	Safety and sharing the road	Personalized learning at school	A judge's decision in case of uncertainty	Preventive and socially responsible management of the transformations
AI learning types	Clustering data into homogenous groups through unsupervised learning	Algorithms of self-driving cars for vision, decision-making (supervised learning and through reinforcement)	Supervised teaching (student concentration) and through reinforcement (homework follow-up policies)	Supervised teaching of past cases of recidivism	All AI from the moment they involve transformations in companies and administrations
Ethical and societal issues (examples)	Privacy: Data confidentiality	Justice: the equitable sharing of public spaces	Privacy: the confidentiality of student data	Autonomy and critical knowledge in decision-making	Justice: the equitable sharing of productivity gains

THEME 1: **PREDICTIVE HEALTH**

INITIAL SCENARIO: **DIGITAL TWINS**

MARCH 10, 2025. Olivier receives a notification on his phone alerting him 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 Oliver'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 with that of the entire population in the "world health cloud", overseen by the World Health Organization since 2023, that 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 zone in Montreal where circulation is organized to give priority to self-driving electric vehicles.

The self-driving vehicles, privately owned or car share (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 allows users to consider doing activities in their vehicle without being disturbed by jerking movements, for example, working, writing, or listening to music. Vehicles with drivers must adapt to these speeds, under penalty of 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 crossing these roads in improvised fashion, safety barriers have also been installed along these roads.

Samia, 30, lives in Rosemont. She's a massage therapist, strongly geared towards therapeutic relationships and an animal rights activist. She lives with her partner, Robin, 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 Montreal pilot zone, the cars are set in "altruistic" mode, which means they preserve the interests of the greatest number of people, even at the expense of the person in the car.

But since this summer, a group of cyclists is tired of seeing the many safety barriers that confiscate public space for self-driving cars. Since the end of August, they have been protesting by organizing "free bike parades" on the borough's boulevards in the name 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 case 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.

THEME 3: **PREDICTIVE EDUCATION**

INITIAL SCENARIO: **ALTEREGO, AI TO HELP WITH LEARNING AT SCHOOL**

AUGUST 28, 2025. Carmen starts her 3rd year as a teacher at the Thérèse-Casgrain Elementary School. Just 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 in this pilot school to improve support for exceptional students and to personalize teaching techniques to different learning styles and needs. Last year, Carmen spotted Samuel's learning disabilities a bit late in the school term. Samuel would struggle with attention, chat with his peers instead of listening and sometimes show aggressive behaviour 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.

This year everything was going to change thanks to AlterEgo, an artificial intelligence that assists teachers. AlterEgo measures in real time the degree of attention 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 the child and when he or she starts to lose focus on the lesson on the work. The device is also able to analyze the variations of reading speed 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 make adaptations to 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 dubitative 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 concludes her presentation with a joke: "When I have to work at night, I'll definitely need AlterEgo to take care of Lola and Emiliano. Maybe one day it will be so!"

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, who would send his grocery list to a partner store, would format 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 moment 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 get to 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 cameras 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 a great number of similar cases.

Following the decision rendered by the algorithm, the judge had the choice of following it or ordering probation for Sylvia, on the condition she follows the all-new rehabilitation program for delinquents but that has no data history, which means no possible interpretation by the algorithm.

The judge, who is favourable to social innovation, chose the second option. The rehabilitation program plans for Sylvia to follow an evaluation and regular individualized control for two and a half years, as well as find a legal occupation. Seeing 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 Montreal, Zéro Carbone Logistique (ZCL) is a new world leader in sustainable logistics, and has seen incredible growth over the past five years. The company currently employs 3000 people in Montreal.

From its launch, ZCL wished to include its environmental and social objectives in its shareholder agreement by adhering to B Corp status[23] and by following the ISO 26,000 standard recommendations on a company's social responsibility. 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 evolve towards types of cooperation between humans and

[23] A certification issued to companies that satisfy societal, environmental, governance and public transparency requirements.

co-bots! For ZCL management, there's no way this evolution is done in brutal fashion, and they wish to establish a "socially responsible restructuring", by carefully preparing the collaborators for new positions.

Nabila, one of the founders of ZCL, suggests the following solution: creating, in partnership with one of the giants of the web, a massive data processing platform used by AI applications in logistics AI. Jean-Raymond, the company's union representative, is very worried: he mentions that these companies feed off of underpaid workers who spend 15 hours a day coding data to train algorithms, and that it is not a respectable 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 player from the Web is ready to invest immediately in massive data for sustainable logistics and create, with ZCL, a subsidiary in Montreal that could hire most of the 1000 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.

ANNEX 3
Other Forms of Participation

When it was placed online in November 2017, the Declaration's website offered two options to get directly involved in the le co-construction process: with an online survey and by inviting people to submit a memoir. Although the results of these consultations will be further expanded upon in the final report, we can already paint a picture of the preliminary results.

A3.1
THE ONLINE SURVEY

The online survey was made up of 35 questions, articulated around the 7 values of the preliminary version of the *Montreal Declaration*. It was bilingual. A little over 80 people answered questions ranging from "How can AI contribute to well-being?" to "Can an artificial agent such as Tay, Microsoft's 'racist' chatbot, be morally blameable and responsible?" and "What kind of legal decisions can be delegated to AI?" These questions sometimes lead to near-consensus: no, it is not acceptable that an autonomous weapon kill a human being; yes, we must fight against a concentration of wealth and power among a small number of AI enterprises; yes, we should know who our personal data is being sent to and who is using it. But these questions also raise important doubts: can AI really guarantee to respect privacy? Is it acceptable for AI to answer email in your name?

We can also see certain divides, especially in the attitude towards the private sector. Some people (the majority) fear a wrong turn for AI dictated by enterprises searching for profit rather than the common good (self-driving cars fan the flames of certain respondents). Others consider the private sector to be the best guarantee that AI develops in independent fashion if it's not tied to any political programs.

More commonly, it's probably the level of trust in AI (and in the future) that most obviously divides people. While some people seek solutions for the problems raised by AI by making it perform better and adapt to human needs, while others worry about the disruptions it will cause in social and economic life, on the edge of dehumanization.

It is also remarkable that many potential solutions echo concrete recommendations that appeared during the library consultations or the co-construction days. We therefore find a promotion of transparency and the idea of creating an AI ombudsman or a "committee of wise people". Finally, at least one leitmotif comes out of this online consultation: it's that AI must be designed as a tool in the service of humans. As one respondent wrote to sum up many similar interventions, "computer systems are there to assist in the decision-making process, and must continue to remain there".

A3.2 SPONTANEOUS MEMOIRS OR RECOMMENDATIONS

As for memoirs or spontaneous recommendations, 15 were received and are available on the *Montreal Declaration* website. These all went in very diverse directions, which makes it harder to paint a coherent picture. Here are a few reflections or suggestions that stood out:

> A warning against the risk of instrumentalizing and assimilating humans to a simple machine that hides a new totalitarian ideology. "AI must not participate in the temptation of humans shirking responsibility at the expense of technology" (Jean-Claude Ravet, editor-in-chief of the Relations journal).

> A call to include Quebec technology companies in the reflection on AI development (Association québécoise des technologies).

> An elaborate concept of "loyal" AI, meaning each of us possesses our own personalized AI, set to only serve its owner - not the company or the State (G. Wark).

> Four paths to reconcile AI development with the protection of privacy (Commission d'accès à l'information du Québec).

> The necessity for a national or international research and oversight organization (an observatory) which would recommend standards to be respected (John McNally).

> Establishing an independent advisory body made up of experts in AI, ethics and law, as well as citizens (Lise Parent).

> An argument so that sharing information is not the default option as far as the privacy principle is concerned, and a warning against a technocratic, and not truly democratic, wrong turn for the *Montreal Declaration* (Ariane Quintal, Matthew Sample and Eric Racine).

> An argument for a minimalist moral conception of AI (Pierre Musseau-Milesi).

> Finally, the Ordre des ingénieurs du Québec notes that the values of the Declaration are "perfectly compatible with the values of the engineer profession" and suggest many recommendations tied to training and diversity as well as adopting "best practices" such as, for example, the European Union's General Data Protection Regulation (GDPR).

We also wish to thank all the citizens and organizations for their written contribution to enhance the reflection on the responsible development of AI:

> L'Ordre des ingénieurs du Québec

> Bruno Robert

> Jean-Claude Ravet

> L'Association québécoise des technologies

> Annick, Guillaume et Raphaël Hernandez

> G. Wark

> La Commission d'accès à l'information

> John McNally

> L'Université du Québec à Montréal

> Lise Parent

> Ariane Quintal, Matthew Sample et Éric Racine

> Pierre Musseau-Milesi

> Human Aware

> And those who wish to remain anonymous.

All the memoirs are available on our website[25].

[25] www.declarationmontreal-iaresponsable.com/propositions-citoyennes

< >

Montréal Declaration
Responsible AI_

</ >

montrealdeclaration-responsibleai.com