< >
Montréal Declaration
Responsible AI_
</ >

PART 4

FALL 2018 CO-CONSTRUCTION: KEY ACTIVITIES

TABLE OF CONTENTS

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

1. INTRODUCTION

The Montréal Declaration's main co-construction activities were carried out from November 3, 2017 to April 31, 2018. The Declaration team nevertheless continued to work on the project in the fall of 2018, organizing three key activities to mobilize the knowledge of more actors on various important issues in responsible artificial intelligence (AI). A day of co-construction was organized in Paris using the winter 2018 co-construction model. To mobilize the knowledge of stakeholders in the cultural sector, a focus group was also held on issues related to the advent of AI in fields related to art and culture. Lastly, as a bridge between the public consultations held last winter and ongoing research, an activity was carried out with graduate students, simulating the drafting of policy briefs, in partnership with the *Comité intersectoriel étudiant* (CIÉ) of the *Fonds de recherche du Québec* (FRQ).

These activities supported further analysis and the drafting of recommendations on public policies (see Part 6 of the report *Priority projects and their recommendations for responsible AI development*). The following sections provide a recap of the issues identified as well as the main potential solutions developed by the participants in these activities. Some draw on the mechanisms proposed during last winter's co-construction, and support the need for their implementation, while others add new elements to the discussion.

WRITTEN BY

NATHALIE VOARINO, Scientific Coordinator, PhD Candidate in Bioethics, UdeM

CHRISTOPHE ABRASSART, Associate Professor in the School of Design at the Faculty of Planning, UdeM

CAMILLE VÉZY, PhD Candidate in Communication Studies, UdeM

IN COLLABORATION WITH

LOUBNA MEKKI BERRADA, Doctoral student in Neuropsychology, UdeM

VINCENT MAI, Doctoral student in Robotics, UdeM

2. CO-CONSTRUCTION DAY OUTSIDE QUEBEC (PARIS, FRANCE)

This section presents results from a co-construction day held in Paris on October 9, 2018, organized in partnership with the Canadian Embassy in Paris, the Canadian Cultural Centre and the House of Canadian Students. At this event, 26 persons of varied backgrounds were mobilized to examine issues related to responsible AI. The participants were assigned to one of three co-construction tables, with each table addressing a key theme in AI development: Democracy, the Environment and the World of Work.

Chart 1: Profile of Participants in the Co-Construction Day in Paris

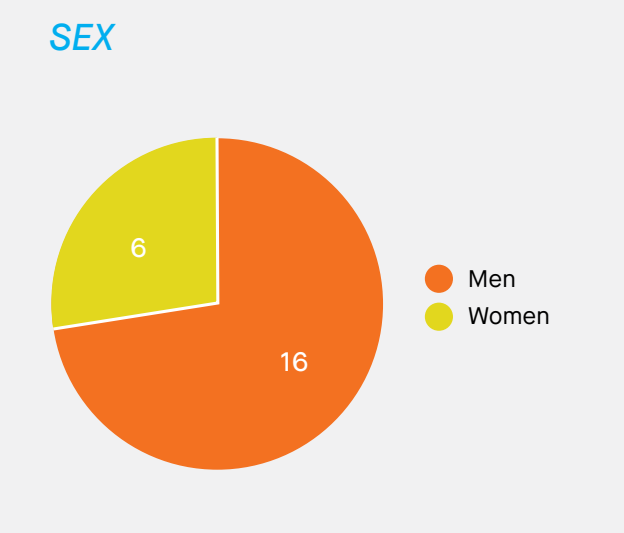


Chart 2: Profile of Participants in the Co-Construction Day in Paris

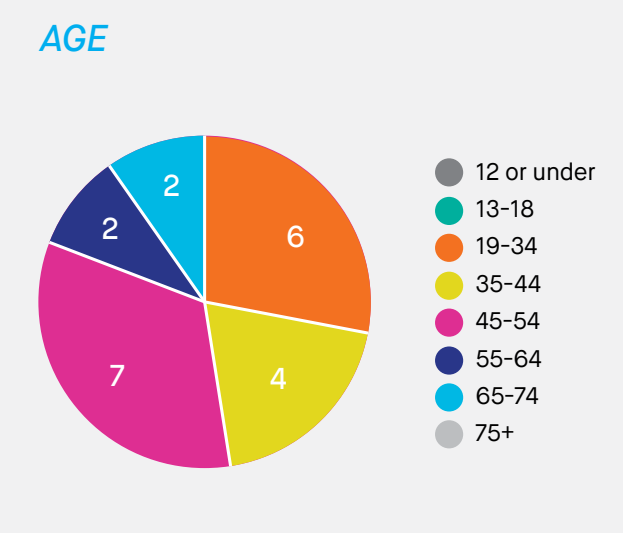


Chart 3: Profile of Participants in the Co-Construction Day in Paris

EDUCATION | (out of 22 respondents) 2 respondents gave no answer

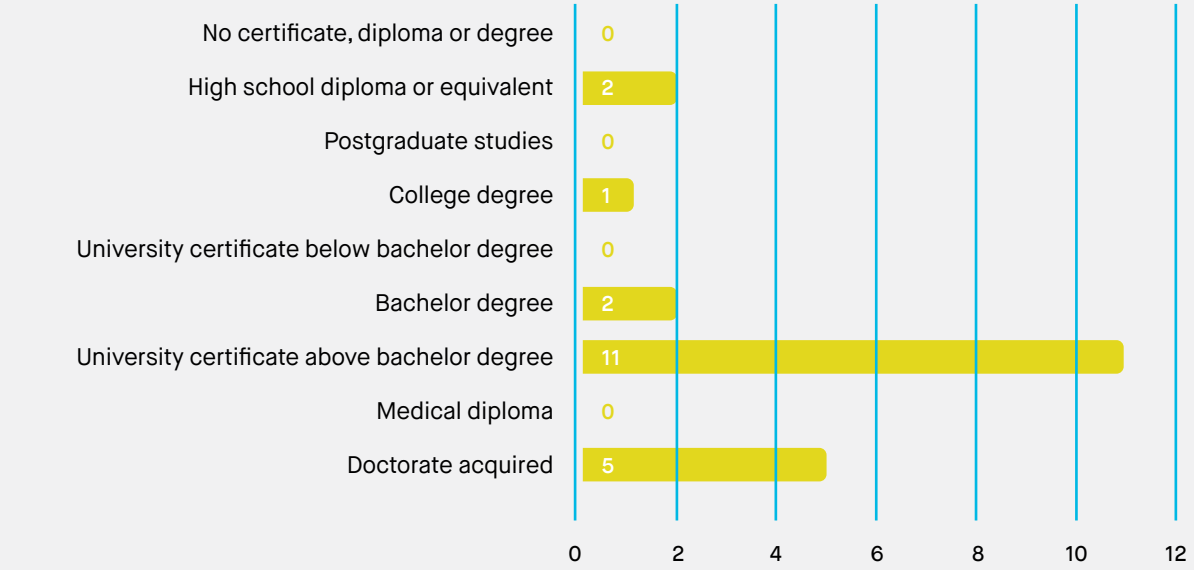
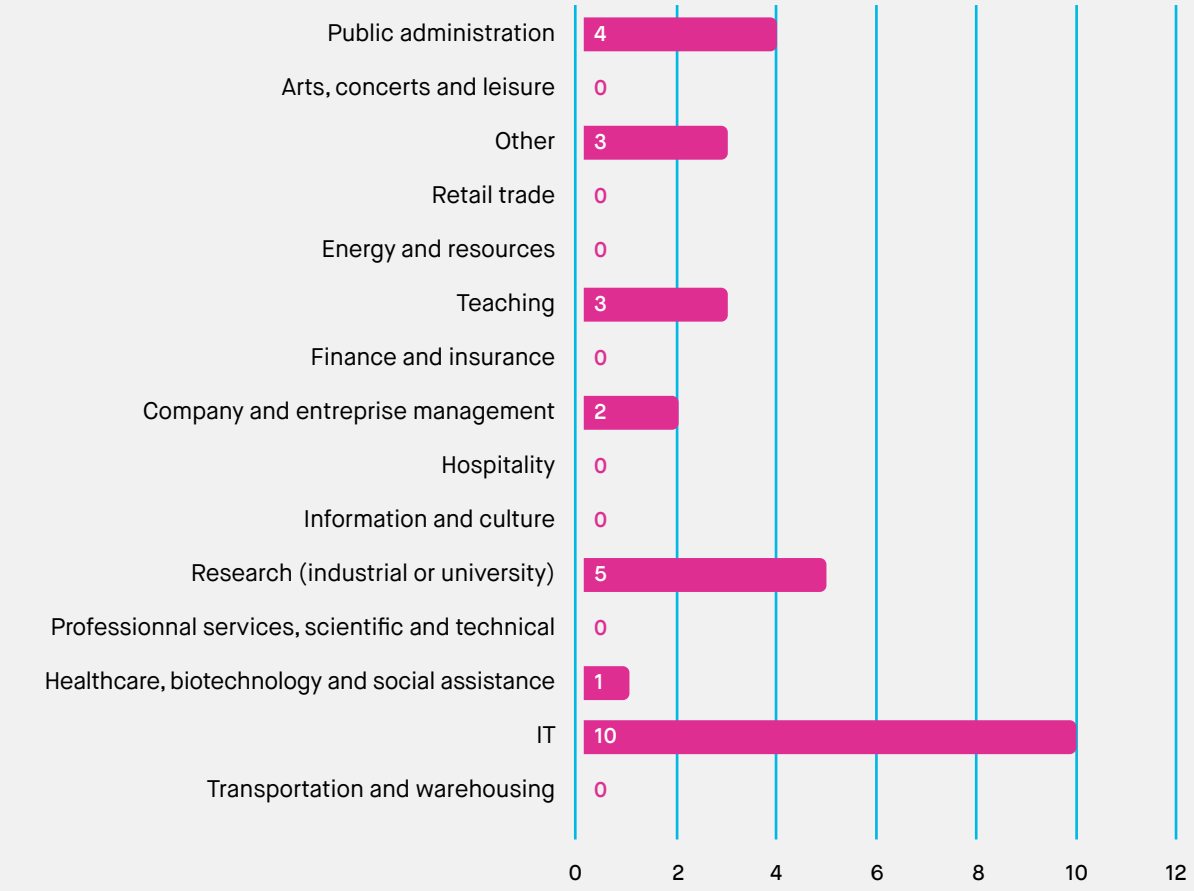


Chart 3: Profile of Participants in the Co-Construction Day in Paris

FIELDS OF ACTIVITY | (out of 22 respondents)

7 respondents indicated more than one sector of activity; 2 respondents gave no answer



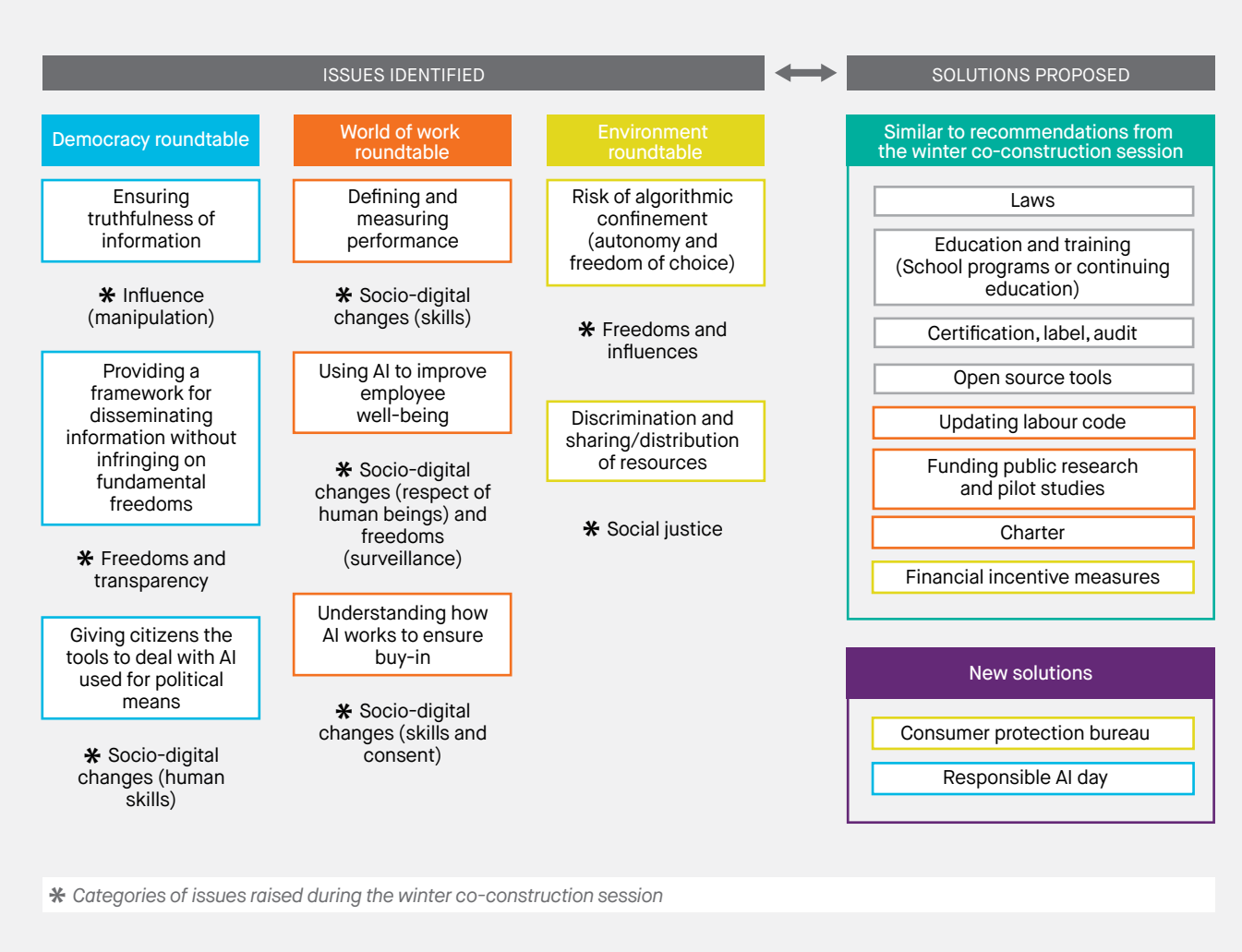
The discussions were organized around three distinct deliberation periods (identifying the issues, preparing recommendations and writing a front-page newspaper article), following the co-construction model developed for last winter's activities. Three trigger scenarios were used (see Appendix 1, in French only), involving the use of AI:

1. to encourage ecological behaviour,
2. in human resource management in a business, and
3. to create fake news during an election campaign.

These three scenarios allowed us to explore issues related to the environment, the world of work and democracy from a new point of view, since we had never used this format before.

The following sections relate the directions taken by the discussions at each of the three discussion tables[1]. They highlight the appearance of issues that had already been identified at last winter's co-construction activities, but in the specific given contexts. In addition, various potential solutions or mechanisms for managing responsible AI were proposed. Some of them are similar to those formulated at last winter's activities (e.g. Laws and Training), while others are new (e.g. Responsible AI Day).

Diagram 1: Issues Raised and Potential Solutions Proposed at the Co-Construction Day in Paris



[1] The citations were taken from post-its written by the participants.

2.1.

ADDRESSING DEMOCRACY ISSUES RAISED BY FAKE NEWS

Summary of the initial scenario

Fake News During an Election Campaign

Two weeks before a presidential election, an emergency meeting is called by the Information Integrity Agency (IIA), which was implemented under the *Act to Combat the Manipulation of Information*. A video has emerged showing the outgoing president making compromising remarks on immigrant workers, and it has gone viral. The president's spokesperson says that the video is fake, created by a foreign agency trying to interfere with the elections, using GAN algorithms (Generative Adversarial Network). Even though dissemination of the video is prohibited, it continues to circulate on various foreign websites. With just one month to go before the first round of the presidential elections, the IIA must come up with a plan on how to contain the devastating impact of this misinformation and restore conditions for a healthy campaign.

The objective of this scenario was to stimulate a discussion of ethical issues related to information manipulation, which can harm democracies when it spreads virally. This is particularly true when artificial intelligence techniques are used to imitate people and change content while maintaining a very high level of realism, making the detection of what is fake very difficult.

The deliberations presented here are the result of a full day of discussions among seven researchers, experts and students working in the fields of ethics, organizational development, machine learning, the social web and political science. Taking this scenario in 2022 as a point of departure, the discussions led to drafting a headline and lead for a front-page article in the responsible AI newspaper, dated October 9, 2022: "First Responsible AI Citizen Day."

First Discussion Period: FORMULATION OF THE ETHICAL ISSUES IN 2022

DEMOCRACY

The participants believe that AI itself is not the cause of the attacks on democracy. These problems already existed; however, they are altered or aggravated by the opportunities afforded by AI. Consequently, we must prepare for this new reality, and make the necessary adjustments. It was therefore suggested that "maintaining a healthy democracy" is a particularly important challenge, in a context where "choices based on wrong information" cause problems. A legal framework and training in critical thinking were discussed as two ways to defend against the impacts of manipulated information and maintain a healthy democracy.

LEGAL FRAMEWORK

The participants wondered how practices related to the manipulation of information could be controlled:

> "Can the law effectively control all AI practices (manipulation of information)?"
> - A participant

In this context, one question came up several times: "how can justice be brought to bear" when one's reputation has been sullied by fake news? One of the impediments is that it is difficult to control the manipulation of information without limiting freedom of expression and other fundamental freedoms.

KNOWLEDGE AND CRITICAL THINKING

"The importance of the interdependency of democracy and education (in particular, education in critical thinking)" was mentioned in a call to develop critical thinking for *"informed participation in public life."* One participant said that instead of mounting a defence against propaganda, it would be better to develop critical thinking, which allows individuals to defend themselves against the impacts of propaganda (indoctrination, changes in choices and behaviour, extreme polarization, etc.). Preventing all propaganda, i.e. all actions taken to influence opinion, could lead to censorship or curtail freedom of expression. Rather we should focus on educating citizens to develop "critical thinking" and "media and statistical literacy."

The issue of **democratizing access to information and knowledge** was then raised. Several conditions must be met: net neutrality must be ensured, in order to guarantee free access to all information, and everyone must have access to technological tools, so that they can obtain information and express themselves. Everyone, without discrimination, must therefore be able to learn about how AI works and its related issues.

AUTHENTICATION AND AI'S ABILITY TO IDENTIFY WHAT IS FAKE

By learning certain AI techniques, some may become able, for example, to develop tools for detecting and correcting fake news. However, this depends on the potential and the need to use AI to sift out the fake from the real in a context where, for example, it is impossible to distinguish a real video from one created by AI with the naked eye. The participants therefore examined "mechanisms for authenticating the news/information" and "technology's ability to understand sarcasm." One example given of a potential authentication solution was the "traceability of sources on dissemination tools (e.g. WhatsApp)." It was nevertheless mentioned that an AI that can automatically censor publications it flags as fake, malicious or unreliable would curtail freedom of expression.

RESPONSIBILITY

Given the impression that "control over information" is lost when fake news spreads very quickly or even goes viral, participants discussed the issue of responsibility in several different stages: creating, disseminating, sharing and reading information. They wondered *when* the truthfulness of the news should be evaluated (before it is created, before it is published, before it is shared, or when it is read?), *by which entity* (an international organization, the state, journalists, readers, the dissemination platform used to publish the news and/or sharing on platforms such as WhatsApp and Facebook?), and *how* (developing the reader's critical thinking, creating a label indicating the information's level of truthfulness and the source's reliability?).

The creation and dissemination of fake news are also seen as combining the roles of several of the actors needed to take on the responsibilities of creating information, disseminating it, determining its credibility, and confirming that its sources are truthful and reliable; i.e. journalists, readers, the entity that should verify the truthfulness of the news, the information dissemination platform and any individual independently creating or disseminating information.

Participants discussed the "ethical conduct of the media (capturing and disseminating information)" and "journalism's credibility." A new role for journalists and the media could be to judge the information published and shared, checking it and declaring whether it comes from a reliable source.

Table 1: Democracy, First Deliberation Period: Identification of Ethical Issues in 2022

Ethical issues in 2022	1	2	3
Description	Ensure that the information is truthful	Manage the dissemination of information without infringing on fundamental freedoms	Equip citizens to deal with the political uses of AI
Related principles	Democracy, responsibility	Democracy, autonomy	Democracy, knowledge, autonomy

Following this discussion, the participants formulated three priorities for which potential frameworks need to be proposed to keep democracy healthy, despite the impacts of propaganda in the form of the automated creation and dissemination of misinformation. Those three priorities are:

1. To ensure that the information is truthful and reliable, in particular to preserve the health of democratic discussion.
2. To control the dissemination of information without infringing on fundamental freedoms, including freedom of expression, in particular through the development of journalistic and technological standards on the dissemination of information.
3. To equip citizens to understand the political uses of AI so that they can learn about them and freely develop their own opinions.

Second Discussion Period: PROPOSALS FOR OVERSIGHT OF AI

In response to these issues, the team's discussions led to a series of five proposals on AI oversight:

Table 2: Democracy, Second Deliberation Period: AI Management proposals for 2018-2020

Management proposals	1	2	3	4	5
Description	Creation of an authority that would certify journalistic standards with a label (red, yellow or green label)	Creation of open-source tools that can distinguish between true and false (e.g. mobile phone app)	A bill to guarantee net neutrality	Implementation of school and continuing education programs (such as MOOC) for developing critical thinking skills	Implementation of a Responsible AI Citizen Day
Related issue	Ensuring the truthfulness of information	Control the dissemination of information without infringing on fundamental freedoms		Equip citizens to understand the political uses of AI	

In line with the previous ideas, the discussions that led to these proposals identified the technical, political and educational dimensions of the problem of manipulated information. In response, the participants proposed a series of measures intended to structure fact-checking of the news (1, 2) and educate citizens for free and informed participation in democratic life (3, 4, 5). This will require standardizing information creation and dissemination practices (certification) and fake news detection support (tools), equal information access for all through a neutral Internet network (legislation), education for all—at every stage of life—to develop critical thinking (school and continuing education programs), and raising awareness about the political uses of AI (responsible AI day).

Brainstorming on the responsibilities of journalists resulted in a proposal for a new certification authority that would establish journalistic standards through a system that includes an information reliability indicator. This would serve as a way to establish the reliability of sources and the credibility of information. The participants felt that such a certification organization should be independent of government.

Instead of a legal framework, the participants preferred an approach based on professional standards and the establishment of guidelines on how to create, fact-check and disseminate information. One participant pointed out that a minimalist approach would be more reasonable for the design of a system that indicates reliability of information, to avoid a situation where the system has excessive influence over electoral processes and other competitive situations in which the news plays a crucial role. For example, this might happen if the reliability indicator system could give one party an advantage, or be used in a strategy to manipulate or lobby against an adversary. At the same time, a minimalist approach would be better placed to avoid excessive limits that could infringe on fundamental freedoms.

In opting for this preventive solution, the participants did not propose any solutions on how to respond to the dissemination of fake news. They nevertheless mentioned that when this occurs, there would be a need to intervene in its dissemination as quickly as possible, and refute the fake news with the fact-checked version.

The fact-checking journalistic practices adopted by this certification authority could employ an open-source tool whose technology would find fake elements even when they are undetectable to the naked eye, in particular when they have been created using AI techniques. The practices should also check to ensure that certain accurate information is not identified as fake news.

The participants also pointed out the large share of the responsibility that needs to be shouldered by information-sharing platforms such as social media (Facebook, Twitter, YouTube, Snapchat, etc.) and instant messaging platforms (e.g. WhatsApp). They wondered about the need to oblige these platforms to install a tool that uses AI to identify fake news. The participants did not appear to trust private businesses to install a fair system for identifying and blocking fake news, given the fact that, for now at least, this type of platform does not flag fake news videos that are influencing the opinions of various audiences. However, it would be a positive step if these platforms had a system that would indicate the level of reliability of a news item, since they absolutely must be held accountable for the role they play in disseminating manipulated information affecting democracies.

Training citizens and raising their awareness are subject to a second series of measures that target the development of critical thinking among citizens, in particular through media literacy that will equip them to freely browse the information universe in an informed manner. This could be achieved through school curricula and public spaces available to all, through public universities in libraries or cafés, for example, or even through creative awareness-raising campaigns that reach people in their day-to-day lives to keep them vigilant about information manipulation practices.

Third Discussion Period: WRITING A HEADLINE AND LEAD FOR A FRONT-PAGE NEWSPAPER ARTICLE FOR 2020

These proposals were then made into a narrative in a headline and lead of the responsible AI newspaper for October 9, 2022, as follows:

First Responsible AI Citizen Day

Several events were held simultaneously in Canada and France as part of the first Responsible AI Citizen Day, particularly to inform citizens about the new opportunities offered by AI and the importance of being able to think critically and to equip them accordingly. To this end, we are very proud to inform you that our paper has been accredited by the Order of Responsible AI Newspapers.

By promoting a "Responsible AI Citizen Day," the participants underscored the importance of raising awareness among all citizens of the need for them to appropriate AI issues to be able to participate in democratic life. The special mention about the paper being accredited as a responsible AI newspaper also reinforces the role played by the media and journalists as important actors in healthy democracies.

2.2. DISCUSSING ENVIRONMENT-RELATED ISSUES

Summary of the 2025 scenario given as a point of departure

The mercury keeps rising, with record-breaking temperatures recorded all over the world. In response to the climate change crisis, various cities are talking about introducing **EcoFit**, a highly incentive-based individual carbon permit system that is connected to the citizen's bank account and various online shopping apps. In these cities, the prices of goods and services are posted in euros and carbon, and each citizen must aim for a total personal consumption representing a maximum of 4 tonnes of carbon emissions per year. This rating gives them access to a series of environmentally responsible transportation, education, professional development and cultural services. In 2025, Ive and Charles have managed to gradually adjust their consumption to meet this target, and even get below it. And since they have spent less, they have saved more than expected. So they are considering a Christmas vacation in Cuba, and have begun visiting travel websites. But then they receive a message on their phone: "Beware the rebound effect: spending your savings on a flight will negate all your hard work. Think about planning a trip closer to home!"

The objective of this scenario was to encourage a discussion of what could be achieved on ethical issues through predictive management (using AIS) of environmental rebound effects on the consumer and equipment markets. Rebound effects can be explained as follows. As equipment becomes increasingly energy efficient and the environmental footprint of consumer goods becomes smaller

through better eco-design, we tend to consume proportionally more equipment, goods and services. This means that our gains are lost rather than locked in. For example, we buy larger screens, fill our homes with more equipment, travel farther in our cars, travel by air, etc. The result is an increase in GHG emissions and even greater pressure on resources and biodiversity. Given these rebound effects, economic development must be considered alongside its material reality and ecological footprint.

The algorithmic apparatus imagined here is part of an "AI for Earth" approach to using AIS. This exploratory scenario also includes predictive and tailored management of rebound effects, through supervised learning based on consumption histories (e.g. bank transaction histories) that are associated with nudges.

The deliberations presented here are the result of a three-hour round-table discussion involving eight citizens with an interest in new technologies and in environmental and sustainable development issues. The discussions based on this scenario for 2025 led to the formulation of an initiative presented in a headline and lead of the responsible AI newspaper dated October 9, 2020: *"Major Achievement for ConsoM'AI: 1 Million Subscribers in a Week. General Regulation on Opening Data for the Environment."*

How did the group's deliberations lead to this original proposal? What were the pivotal moments in their discussions? How did the ideas develop in each stage? The following sections present some key moments from this group's deliberations along with our comments.

FIRST DELIBERATION PERIOD: FORMULATION OF ETHICAL ISSUES IN 2025

During the event, participants jotted down their questions about the various principles underlying the Montréal Declaration on a series of Post-its:

PRIVACY PRINCIPLE

"Can we reconstruct this family's entire consumption history?" "Will the database be managed reliably enough to protect personal data and win users' trust?" "Could there be a right to erase?"

AUTONOMY AND FREEDOM OF CHOICE PRINCIPLE

Does this AI application lead to a new "prescriptive power?" "Does it maintain autonomy in decision-making and free will?" "How can one think critically about these personalized recommendations?" "Does this represent a machine exercising control over day-to-day life?" "Is there a risk of algorithmic confinement, of algorithmic bubbles?" "How can such a system take into account the singular context of a purchase decision (e.g. an emergency situation)?"

RESPONSIBILITY PRINCIPLE

This measure needs to help "strengthen environmental responsibility in day-to-day life," and "to express personal ethics as an actor in consumption." But, when relying on AI tools, "is there a risk of externalizing personal responsibility?"

JUSTICE AND EQUITY PRINCIPLE

By asking businesses to assess the carbon footprint of their products and services before they enter the market, does this application "ensure free competition?" "Is there a risk that it will expand the market power of large corporations and discriminate against SMEs by creating a barrier to entry, due to the cost of these environmental assessments?"

"How will fair trade, which has other ethical dimensions, be evaluated?" "Will some producers be favoured over others?" "The rich will be able to consume more and buy carbon quotas to offset their emissions. This is social inequality!" In addition, if "everything is reduced to data and the market," "will initiatives to reduce non-market GHG emissions (e.g. a project for people living in neighbourhoods that offered active mobility, in urban agriculture) become invisible and therefore be discriminated against?" Lastly, "Lifestyles differ around the world, diets vary (e.g. vegan, religious). Is there a risk that some lifestyles will be favoured and others will be discriminated against?" "Will this create culture-based discrimination?"

DEMOCRACY AND GOVERNANCE PRINCIPLE

"Who will regulate this system? The United Nations? The rich countries? How will abuses be monitored?" "Should an authority be established to regulate carbon footprints?" "If some CO_2 is saved, could those savings be transferred to family and friends?" "Should the recommendations that will be given priority be subject to public debate?"

The participants then engaged in several in-depth discussions, going back to their initial ideas to generate more ideas. Then, following close to 45 minutes of discussion, they selected their priority groups of ethical issues for 2025 by applying coloured labels. Two principles of the Montréal Declaration emerged: Autonomy, tied to freedom of choice; and Justice, which they associated with the Equity principle.

Table 3: Environment. First Deliberation Period: Formulation of Ethical Issues in 2025

Ethical issues in 2025	**1**	**2**
Description	Risk of algorithmic confinement due to this new prescriptive power and the configuration individuals' preferred spaces. How can individual and societal autonomy be maintained? How can we account for an initiative to reduce carbon emissions that is outside the system?	Carbon offsetting could favour the richest members of society. What limits should be assigned to them? Conversely, could people who consume only small amounts of carbon redistribute the carbon they saved? What about relations between Northern and Southern countries? Is there a risk of cultural discrimination?
Related principle	Autonomy and freedom of choice	Justice and equity

This selection of priority issues by the team led them to develop and give more clarity to two ethical issues in AIS. The first is related to the Autonomy principle, with potential actions to reduce non-market greenhouse gas emissions (e.g. a citizen initiative on daily mobility). The second issue concerns the Justice principle, with potential carbon offsetting for the richest members of society, or emission sharing for citizens who consume less than their limit.

SECOND DELIBERATION PERIOD: PROPOSALS FOR MANAGING AI FOR 2018-2020

In response to these issues, the team continued their discussions, brainstorming on the four related principles. The participants formulated several proposals for managing AI. Three of them are presented here, illustrating how the ideas were developed into a headline and lead for a front-page newspaper article.

Table 4: Environment, Second Deliberation Period: AI Management Proposals for 2018-2020

Management proposals in 2018-2020	1	2	3
Description	• Develop a code of ethics for system designers, programmers and managers (e.g. to ensure that the prescriptions support equality).	• Create a consumer ombudsman, an independent administrative authority, that is audited by the national democratic assembly. • Audits of the system, of the diversity of choices and recommendations, and the publication of transparent reports. • Support citizens in their autonomy.	• Provide substantive financial support to help people with the most modest means to adjust. • Allow people to exceed the annual target, but with a growing marginal cost for each additional tonne of carbon.
Instrument categories	Laws and regulations Code of ethics	Institutional actor	Incentives and support measures

These proposals, which reflect true institutional creativity (that went well beyond the examples of very general tools provided in the participant's booklet), are in keeping with the issues identified in the previous step. The proposal to create a consumer ombudsman to regularly evaluate the system by performing audits, being publicly accountable, and organizing citizen support also shows a further development of the ideas formulated in the previous step. It is on the basis of this proposal that the participants developed their headline newspaper article in the following step.

THIRD DELIBERATION PERIOD: WRITING A HEADLINE AND LEAD FOR A FRONT-PAGE NEWSPAPER ARTICLE FOR 2020

These measures were then expressed in a poster as follows. The team developed the following headline and lead for a newspaper article dated October 9, 2020:

Major Achievement for ConsumAI! One million subscribers in a week.

General Regulation on Opening Data for the Environment.

Following the passage of GRODE (the General Regulation on Opening Data for the Environment), which required personal data to be made public, the CONSUM'AI organization conducted a major survey on the freedom of choice of ECOFIT users and found many limitations and algorithmic bubbles. The first recommendation in the CONSUM'AI report is to develop the means for counter-expertise, and train of users to ensure true pluralism and everyone's participation in reducing greenhouse gas emissions.

So if the use of AI presents a certain potential for managing the environmental issues associated with consumer behaviour, this perspective also raises many ethical issues that must be properly framed.

2.3.

ADDRESSING ISSUES OF DIGITAL TRANSFORMATION IN THE WORKPLACE

Summary of the initial scenario:

Mining HR data to optimize work atmosphere

Peter has finally landed a job in a good law firm. After his first three weeks on the job, he meets with Marco from Human Resources for some personal mentoring. Marco explains that, going forward, the firm will be using AmbAI+, a conversational analysis AI that studies employees' attitudes and helps maintain a peaceful and productive atmosphere at work (the system analyzes all e-mails, telephone calls and conversations in work meetings). AmbIA+ provides personalized assistance, advice and training, but no disciplinary action is taken. Peter is told that all his conversations in the office have been monitored, except for those on October 15 and 16. "On several occasions you interrupted your co-workers in meetings to repeat the same ideas, and this has created some tensions. Apparently, the algorithm also detected periods of inactivity on the network, periods lasting several hours, when you had engaged in no conversation with your colleagues. This does not pose a problem in and of itself, but it is better to stay in touch with your team. Do you remember why you were inactive on the network at these times?" This not only has Peter worried, it also embarrasses him, and he wonders how these issues can be relevant.

The objective of this scenario was to open a discussion about the ethical issues related to businesses using AI to monitor and manage their employees. The AmbAI+ system imagined here is used to optimize performance and control the work atmosphere using data mining techniques.

The deliberations presented here are the result of a full day of discussions among a group of 10 engineers, AI designers, digital strategy managers, investigators, students and academics. Based on this scenario for 2025, the discussions led writing a headline and lead for a front-page article in the AI newspaper dated October 9, 2025: "First Employee Terminated Because of AI."

FIRST DELIBERATION PERIOD: FORMULATION OF ETHICAL AND SOCIAL ISSUES IN 2025

In this first part, the participants identified five categories of issues related to the development of AI applications in the workplace.

AUTONOMY

First, the discussions underscored the issue of respect for autonomy (in particular as it relates to employees' ability to act). The participants criticized a kind of manipulation of "how people feel things," and a "forced" organizational culture. They were troubled by how information was being saved, such as information on employees' behaviour and interactions with each other, for the purposes of cultivating an organizational culture. In operating this way, the company is somehow trying to standardize employees, which could lead to a great deal of tension (even "totalitarianism," if everything began to be measured in order for the business to exercise control over its employees), through insistent recommendations issued by the AI. Respect for autonomy is tied to respect for employees exercising a certain "free will" as well as respect for their emotions, which in this case has not been given due consideration, according to the citizens.

Should all this data be kept (the smallest actions are being observed) and used for this purpose? The participants therefore raised an issue related to "surveillance," which could limit the scope of employees' actions and speech (this is closely related to the issue of respect for privacy).

> "AI is watching"
> - A participant

The citizens put forward the need to foster autonomy by allowing everyone to work in whatever way is best for them, in the best interests of the business. Employees should be able to have control over their data, the employer should tell employees what data is being collected and use this data in a carefully circumscribed manner, and everyone should be free to "disconnect," especially to maintain a boundary between what is professional and what is private.

PRIVACY

The participants debated where this boundary between the private and the professional in a business is situated, and concluded that it is difficult to define. Some of them felt that the AIS in question would intrude on employees' privacy, based on their conversations:

> "When should discussions be considered personal?"
> - A participant

On the other hand, some participants mentioned that, ordinarily, anything related to one's private life should not be discussed in e-mails or telephone calls using company equipment (meaning that it would not be analyzed by the AIS presented in this scenario). These participants asked: Is there a place for one's private life in a business?

A consensus nevertheless emerged about the use of AI: as a business tool, it must never, under any circumstances, be used to analyze anything private. The issue then becomes how this boundary should be defined, in order to clearly identify when AI can and cannot be used (i.e. there is a need to define what data is purely work-related and can therefore be used in the system's analyses).

The behavioural analysis performed by the AIS was also criticized. It could infringe on both privacy and autonomy (meaning that the citizens then questioned the ethics of using an AIS to track conversations and behaviour, even if they were work-related). Here they criticized a form of intrusion and breach of confidentiality that would result from this constant surveillance:

> "It's as if Big Brother is watching."
> - A participant

Some of the participants noted that these issues were not specific to AI, while others believe that the high level of traceability afforded by AI enhances the relevance of this issue.

WELL-BEING

The participants felt that this type of system could have both positive and negative impacts on employee well-being. While this technology can be used to help employees and improve the quality of their work life (by helping improve relationships or revealing incidents of harassment or intimidation, or even by helping prevent suicide), it appears that the technology can also cause harm ("destabilizing" employees, making them "distraught," standardizing employee behaviour). The participants agreed that the scenario asks too much of employees, who should not be obliged to justify all of their actions. Here the citizens highlight a dilemma between employee well-being and freedom (just how far can a business go in monitoring employees' actions in the interests of protecting their well-being without unduly restricting their freedom to act? When can surveillance be considered "well used"?). The citizens also found a correlation between well-being and the performance objectives: a happy employee is also more productive. So protecting an employee's well-being appears to be good for both the business and the individual.

TRANSPARENCY

The citizens began by discussing the lack of respect for employee consent, since the employee did not know that he was "under surveillance," and they called for more transparency from the employer, who should have informed him about what was and was not being recorded (in particular, through the business's employee training).

> "Employee consent is important in data collection. Transparency with employees is essential."
> - A participant

This transparency issue raised several questions: What are the rules on relationships within a business? What hierarchy has been established for the importance of the data collected? Who has access to it? Do employees have the right to see their boss's data?

The transparency issue also refers to AI or how it can be made interpretable. Here the citizens spoke of the need to communicate about how the algorithm works, including so that individuals will "buy in" to these new systems. They also mentioned the importance of not making a decision based on the conclusion of an AI technology that cannot be explained.

PRODUCTIVITY AND PERFORMANCE

The issue of employee performance and how it is evaluated was then raised several different times. To what extent does AI need to intervene in the interests of improving employee productivity (e.g. by stopping employees from repeating themselves in group discussions)? Is this truly important? Here it would be necessary to define exactly how the AIS can measure and improve performance. There is a risk of emphasizing productivity at the employee's expense, at the expense of his or her personal development and behaviour in the workplace. Does an IA application like this truly give employees an opportunity to improve? Should expectations and objectives be tailored to the individual?

"AI doesn't forget anyone."
- A participant

On the other hand, the participants were concerned with the very definition of what should be considered productive (and counter-productive), as essential to the debate. Which indicators should be used? How are they relevant? Who can and should determine them? Do they need to be related to the business's objectives? Should these objectives be reviewed, based on the AIS's analyses? Even though the system is being used to confirm that employees are "performing," by interpreting the employees' results the AIS could just as well impede innovation, in particular since certain tasks are easier to measure than others.

All the discussions about defining performance, productivity and respect for employee well-being led to the conclusion that these issues are closely tied to corporate culture, which can vary widely from one business to the next and reflect various objectives, interests and values.

Some citizens around the table were concerned that the adoption of AI-based tools is pushing companies to impose notions of performance that are systematically associated with a score, which could result in a form of standardization. Others pointed out that these practices already exist, and will only be amplified or even "industrialized" by AIS, which can process more information, much more quickly. Others mentioned that it may nevertheless be useful to standardize practices, in particular as a response to business needs.

A debate then ensued on the so-called objectivity of AI, raising questions such as: How can we know the decision-making and automation criteria? How will AI define an absence of employee activity? How can we balance performance, objectivity and standardization by AI technologies, on the one hand, with human subjectivity, specificity and arbitrage, on the other?

After discussing these issues for more than an hour, the participants selected three of the five issues as being, for all intents and purposes, priorities for 2025. In some ways these issues reflect principles set forth in the preliminary version of the Declaration.

Table 5: Priority Issues

Ethical issues in 2025	**1**	**2**	**3**
Description	How can we introduce a performance measurement framework while respecting both the goals of the company (productivity) and individual (normalization)?	How can we ensure that companies and employees understand AI (and ensure their buy-in)?	How can AI ensure (contribute to, support) employee well-being?
Related principles	Performance	Transparency (knowledge)	Well-being

SECOND DELIBERATION PERIOD: PROPOSAL FOR MANAGING AI, 2018-2020

For this second part of the activity, participants were asked to formulate recommendations and imagine what kinds of solutions could be implemented in response to these three issues. For each of the issues identified, the participants formulated a wide range of more or less restrictive mechanisms, and ultimately selected six principal mechanisms to implement that would cover all the issues.

In response to the **performance** issue, the participants first recommended organizing **continuous training activities** in businesses to support people in each stage of their company's "digital transformation." Digital **education** was proposed to encourage the establishment of a learning climate, but also to reduce the fear that can come with implementing AI in the workplace.

> "Continuous training for all employees at each stage of a business's digital transformation (encourage a continuous learning climate)."
> - A participant

In addition, participants proposed establishing an **independent administrative authority (IAA)** and a **correspondent** in the company as potential solutions, in particular in order to guarantee respect for the GDPR[2] (which should be extended to the traceability of data and the explainability of algorithmic decisions). The correspondent would be charged with applying the rules and supporting the complainant when a problem arises, and could seek the assistance of the IAA if required. The participants also proposed creating **indicators** that are directly tied to not only the business's objectives (e.g. financial results) but also to certain "human values" (e.g. the employee's well-being). To this end, the citizens felt that it is absolutely essential to pass a **law**, or else this score would be solely correlated with the business's financial interests.

A recommendation was also made for the government to create a **public research program** on algorithm interpretability and transparency (to close the gap in private research on these issues). The goal would be to understand how algorithms make decisions and limit the monopoly held by large AI companies. For the same reasons, **"pilot groups"** (or "test groups") should be established in order to measure the impacts (including the psychological and sociological impacts) of businesses using AI and confirm its relevance and usefulness.

In response to the ***transparency*** issue, some participants argued for less restrictive measures, such as a **mandatory communication** on the various rules followed, the salary data used, the objectives behind their collection (individual or group scores?), and what can be deduced from the data or the results of the analyses performed by the pilot groups mentioned above. This communication must be addressed to all departments (HR, IT, Marketing, Legal Services) and include AI concepts, with training on what an algorithm is and how it learns from data.

To protect ***well-being,*** the participants recommended **an annual evaluation** of employees' perceptions of the use of AI applications. This evaluation could eventually be consolidated by a committee (such as an occupational health and safety committee) that would be responsible for responding when problems arise. The committee should foster good relations between workers and the employer and ensure that everyone's way of working is respected, even when the technology is applied. A **certification** (or label) should be developed, guaranteeing good ethical, environmental and societal practices, and it should be imposed by the state. This will guarantee that businesses meet minimum criteria in order to optimize productivity and performance. **Indicators** of well-being need to be taken into account, just like indicators of performance.

In response to the well-being and transparency issues, the citizens proposed that a law be introduced to define and impose interpretability of AIS (in particular, justifying the decision and guaranteeing access to explicit rules), and it should set minimum criteria for protecting individual well-being (including a right to regularly disconnect)

[2] General Data Protection Regulation, the European regulation that took effect on May 25, 2018.

in order to guarantee the protection of fundamental rights that could be threatened by the development of AI.

For all these issues and in the interests of "regulating without penalizing," an official **charter**[3] would be created, covering the rights, duties and values that should be defended to protect individuals in a company.

In the end, the citizens reached a consensus on **6 recommendations** that covered the major points of the previous proposals:

Table 6: Proposals Retained

Management proposals	**1**	**2**	**3**	**4**	**5**	**6**
Description	Law that defines and imposes AI interpretability and establishes minimum criteria to protect individual well-being	Certification (or label)	Training for different company stakeholders	Updating the labour code to reflect the new digital reality	Creating a charter of rights and duties	Funding public research and pilot studies on AI and its impact on the workplace
Related issues	*The six (6) recommendations were formulated in response to the three (3) priority issues*					

[3] The participants pointed out that a charter is not as limiting in France as it is in Quebec.

THIRD DELIBERATION PERIOD: WRITING A HEADLINE AND LEAD FOR A FRONT-PAGE NEWSPAPER ARTICLE IN 2020

This step involved storyboarding one of the solutions proposed in 2020. Here the participants outlined the risks of using AI in businesses and one of the planned measures for addressing these risks.

First Employee Terminated Because of AI

An employee was terminated after three weeks of work on a recommendation made by an AI application. The employee appealed his dismissal with the labour board and a decision was handed down following a debate in the legislative assembly. A law that will provide a legal framework for this practice will be put to a vote.

The participants mentioned that in their discussions they paid particular attention to the potentially *negative* impacts of AI in the workplace. However, they recognized that there could also be many benefits to using AI, for both employees and businesses, and that these benefits could be addressed in another forum. Introducing an internal body supported by a legislative mechanism would appear to be indispensable to responsible AI in private-sector organizations.

3. DISCUSSION OF THE THEME OF CULTURE with members of the Coalition for the Diversity of Cultural Expressions (CDCE)

In order to address issues related to AI developments in arts and culture, a discussion workshop was organized with the Coalition for the Diversity of Cultural Expressions (CDCE) on September 25, 2018, bringing together 11 experts and stakeholders in arts and culture. A series of discussions were organized around three themes:

1. copyright,
2. cultural diversity, and
3. propaganda and manipulation.

Following the discussions, the CDCE produced a particularly relevant brief[4] describing various challenges and opportunities related to AI developments in the field of culture. This brief also presents the ethical principles essential to responsible AI in culture and the main recommendations that emerged from the discussions on September 25. This section summarizes the discussions of September 25 and the main points of the CDCE's brief.

3.1.

THREE THEMES PROPOSED BY THE DECLARATION TEAM TO FACILITATE DEBATE ON AI DEVELOPMENT ISSUES IN THE FIELD OF CULTURE

COPYRIGHT IN A CONTEXT OF CO-CREATION BY AIS

The use of AIS to generate works at very low prices (e.g. by generative adversarial networks) in music, in visual arts, TV series or in writing newspaper articles will raise the copyright issue in a novel way. Should the copyright belong to the writers of the examples from which the algorithms learn, or to the programmers of the algorithm, or even proponents of the project? Does the remix produced by a generative algorithm constitute plagiarism? If AI replaces artists, what impact will this have on cultural diversity? And what if an AI application "interprets" a work of art? What access should algorithms be given to our artistic heritage?

This problem is of particular concern in the context of AIS that generate creations ("Applications of AI in the cultural field," CDCE brief, p. 3).

ISSUE OF CULTURAL DIVERSITY WHEN NEW AI APPLICATIONS PRODUCE RECOMMENDATIONS BY ALGORITHM

Given the risks of standardized tastes and behaviours, of an "algorithmic bubble," of recommendation algorithms capturing our attention and formatting our choices, how can we maintain a diverse cultural offering? What sort of free, autonomous and critical reception practices should users be encouraged to use? How can we disconnect (connected objects, domestic robots), given the strategies for attention capture? How will the algorithm's contributions be made transparent and be explained to users? Should a public policy on culture be proposed for this diversity issue? What

[4] For more information on issues related to the development of AI and cultural diversity, see: https://cdec-cdce.org/en/ethical-principles-for-the-development-of-artificial-intelligence-based-on-the-diversity-of-cultural-expressions/

will be the new funding mechanisms for cultural diversity?

Above all, this problem concerns algorithms for data recommendations and use (see "Applications of AI in the cultural field," CDCE brief, p. 3).

ALGORITHMIC CENSORSHIP AND CONTEMPORARY ART

Recognition algorithms are used by social media to exercise censorship in ways that are sometimes considered excessive. Contemporary artists have responded to these applications with interventions intended to both make these rules visible and bypass them. How much critical freedom will tomorrow's contemporary artists have? What kind of training do artists need in order to develop critical knowledge?

Above all, this problem concerns recommendation algorithms and data use (see "Applications of AI in the cultural field," CDCE brief, p. 3).

3.2

PROMOTING CULTURAL DIVERSITY IN THE AI ERA

DISCOVERABILITY AND HOMOGENEITY OF CULTURAL CONTENT

"When recommendations are based, among other things, on the popularity of the content, they contribute significantly to the concentration of listening (0.7% of titles represent 87% of plays on online music services in Canada), favouring a minority of artists."
(CDCE brief, p. 5).

Discoverability was identified as an important issue in this new era of AI in cultural production. If the parameters of AIS are correctly set, they may become tools for cultural diversity by expanding global audiences. They could offer relatively diverse content by, for example, allowing artists to increase their visibility without necessarily having to be supported by intermediaries, even without production costs. On the other hand, the participants were concerned about the risk that cultural content would be standardized and about real net neutrality. For example, francophone content from Quebec is rarely proposed by recommendation algorithms on cultural platforms such as Netflix, Amazon and Spotify. These platforms carry massive content offerings, so much that users often just rely on the recommendation algorithm to make choices for them. The participants mentioned the risk that the algorithms will confine individuals to "a particular taste," preventing them from discovering any other content than what is recommended, based on their previous selections. The risk is that this will lead to homogeneous cultural content, in particular because most artists adapt their creations to this mode of dissemination.

The participants said that the audiovisual sector needs to make an effort to make their works discoverable on digital platforms. The problem

does not only stem from AIS; it is also an issue tied to the industry's objectives. If governments have a responsibility for ensuring cultural diversity, multinationals challenge this power. For example, Canada was the first signatory to the UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions. Even though some content may be funded from public sources, the dominant model remains a business model, which may be a problem, such as when we see that Quebec books are not recommended by the Amazon algorithm (on Amazon.ca), despite a desire for them in the province.

The participants therefore insisted on a democratization of creative power and on using AIS for this purpose, such that they become significant agents of the cultural ecosystem for the emergence of creations.

CONSEQUENCES ON EMPLOYMENT

"In May 2018, researchers released the results of a survey of more than 350 AI researchers. On average, they predict that AI will be able to outperform humans to produce school-level essays in 2026, popular songs in 2028 and best sellers in 2049."
(CDCE brief, p. 4)

The possibility that AI will be used to generate an entire work without any involvement by an artist gives some cause for concern. By whom will these works be developed? Only by businesses with a certain amount of capital? Is this compatible with the development of a society consisting of more artists and more diversity? On this point the participants criticized the risk that there will be a closed loop (in particular in light of work on the European directive on this subject), in which the work of only a tiny minority of artists will be favoured by the algorithms. The participants were afraid that the number of artists will decline precipitously.

How can human artists set themselves apart from AI applications? How will this affect personal and collective capacities to create and innovate? In an experiment conducted by ACTRA (the Alliance of Canadian Cinema, Television and Radio Artists), subjects were unable to distinguish between characters created by an AI application and those created by people. This appears to go well beyond the issues raised by deepfake, since in this case new faces were created. But this is a role played by many of ACTRA's members, who work in the video game industry. The issue here is the job losses that these technologies could engender.

The issue of remuneration was also raised, opening the door to a discussion of copyright. Who will be remunerated, and how, if algorithms are used to create works? The remuneration of actors is calculated in number of days and by residual rights. In terms of residual rights, compromises are to be expected, such as standards governing the re-use of works. The participants also wanted to soften the potential impact of AI on artist recognition, which could be less than anticipated, in much the same way that the digital book did not kill the paper book.

RETHINKING COPYRIGHT

"Obviously, the dematerialization of cultural content, technological changes and the arrival of new players who have transformed business models have a major impact on artists' remuneration and the payment of copyright royalties."
(CDCE brief, p. 6)

Various questions were raised on this subject: Where will the royalties go? Which regulations will govern copyright? How much of a manuscript does an algorithm need to produce another one? How will AI be able to draw from other books to bring new ones to market? The arrival of creative AIS therefore raises major property and cultural content issues. Concerning music, the high processing capabilities and greater amount of available data, combined with major advances in algorithm performance,

has led to the creation of artistic "generation" tools. In music, generally, learning is already based on what was done before. This is what AIS do, but at unprecedented speeds.

The participants recognize that AI is already flouting copyright laws, and they questioned whether this type of right could be granted to a machine. Given that satire and parody are already considered exceptions to the rules on copyright, would it be possible to allow an exception for AI? The Copyright Act is currently under review. In this case the participants felt that it is essential for the law in its new form to include amendments related to AIS developments.

3.3

THE CDCE'S ETHICAL PRINCIPLES

In its brief, the CDCE recommends adopting four ethical principles to guide the development of AI and prevent abuses in the cultural field. According to the CDCE, application of these principles should ensure that AI will more successfully integrate cultural issues in general, and diversity issues in particular. These three principles are consistent with those developed in the final version of the Montréal Declaration, but integrate priorities specific to culture and the arts.

The CDCE put forward a principle on the diversity of cultural expressions, which directly reflects the principle of inclusion and diversity. However, in this case the CDCE specifies that this principle should ensure that AIS:

"- enhance local cultural and linguistic content within the populations from which it originates, thus promoting social cohesion as well as the local economic fabric;

- encourage users to make discoveries outside their environment;

- facilitate the transition between technological families (e.g. Apple), rather than locking them in;

- promote interaction and content sharing."
(CDCE brief, p. 7)

The CDCE also proposes a principle on enhancing culture, artists, creators and producers of cultural content, meaning that AIS should help avoid the current devaluation of cultural content and be prevented from "promoting excessive appropriation of revenues that should be directed to cultural ecosystems" (CDCE brief, p. 7). While this principle is related to other principles of the Declaration, above all it is a call to respect the sixth equity principle: "The development and use of AIS must contribute to the creation of a just and equitable society," and the related sub-principles.

The CDCE then proposes a transparency and dialogue principle (transparency in terms of the algorithm's code but also the data used, and dialogue with users, in particular). This principle calls for respecting the fifth principle of the Declaration regarding democratic participation: "AIS must meet intelligibility, justifiability, and accessibility criteria, and must be subjected to democratic scrutiny, debate, and control"; and the second principle of the Declaration, on autonomy: "AIS must be developed and used while respecting people's autonomy, and with the goal of increasing people's control over their lives and their surroundings."

Lastly, the CDCE proposes a principle on the primacy of the public interest, which it defines as follows: "Not all technological innovations are desirable. The development of AI should always focus on improving the quality of life of the population, social cohesion and democratic practices. Governments must defend the public interest against developments that could have rather negative impacts on society." (CDCE brief, p. 8). Respect for this principle is aligned with respect for first principle of the Declaration, on well-being: "The development and use of artificial intelligence systems (AIS) must permit the growth

of the well-being of all sentient beings," but also the eighth principle of the Declaration, on prudence: "Every person involved in AI development must exercise caution by anticipating, as far as possible, the adverse consequences of AIS use and by taking the appropriate measures to avoid them."

The CDCE's ethical principles	Principles of the Montréal Declaration
Diversity of cultural expressions	Principle 7, Diversity inclusion principle
Enhancement of culture, artists, creators and producers of cultural content principle	Principle 6, Equity principle
Transparency and dialogue principle	Principle 2, Respect for autonomy Principle 5, Democratic participation
Primacy of the public interest principle	Principle 1, Sustainable well-being Principle 8, Prudence

3.5

SELECTED RECOMMENDATIONS

Various recommendations emerged from the September 25 discussions. They were formulated with an eye to promoting Quebec cultural content and making citizens aware of the impacts that AI development can have on culture. First, to foster diverse cultural expression in the digital world, the participants recommend that minimal requirements be set on the representation of Canadian cultural content in the recommendations made by algorithms. This is already the case for Quebec TV and radio. The participants do not believe that free markets will develop these requirements on their own, so they must be formulated in laws and regulations.

During their discussions the participants recognized that literacy in AI development is essential. People need to be equipped to understand where the recommendations made by algorithms will lead them. The participants recommend implementing a user education policy, intended to counter the false impression of choice by encouraging users to vary their browsing and stay vigilant to the influence exercised by algorithms. This policy will take the form of education in how to exercise critical choice. Everyone should develop a form of intellectual self-defence, beginning in childhood. The participants also recommend raising awareness among IT developers of AIS's impact on culture.

These discussions also produced a recommendation concerning the transparency and explainability of algorithmic recommendations. Users should be systematically informed when a recommendation has been made by an AIS, and they should have easy access to explanatory information on both how algorithms work and the existence of other cultural content.

The participants also recommended that the businesses developing AIS that have an impact on culture spend a portion of their sales revenue on promoting cultural diversity, such as by funding certain libraries, cultural events or media. The participants also support the monitoring of taste profiling and the protection of personal information, or even the development of an "AI in culture laboratory" to observe algorithms, learn how to interact with them and, eventually, how to influence their development.

Two of the main recommendations that emerged from these discussions were further developed in the CDCE brief:

1. education and training, and
2. revisions of laws affecting the cultural community.

These recommendations are consistent with those formulated for other sectors during last winter's co-construction: 26% of the recommendations refer to legal provisions and 19% to training *(see Part 3 Summary report of the recommendations from the winter co-construction workshops).*

4. BRIDGING THE GAP BETWEEN THE PUBLIC CONSULTATIONS AND A NEW GENERATION OF RESEARCHERS: POLICY BRIEF SIMULATION

4.1.

DESCRIPTION OF THE ACTIVITY

To bridge the gap between the emerging generation of researchers and citizens, the Declaration organized a simulation in partnership with the Comité intersectoriel étudiant (CIÉ) of the Fonds de recherche du Québec (FRQ) and the École de politique appliquée (EPA) at the Université de Sherbrooke. The simulation was held in association with the Journées de la relève en recherche (J2R) organized by ACFAS. The purpose of the "Policies and Artificial Intelligence" simulation was to bring together students representing a new generation of young researchers to produce three policy briefs on AI. The objective was to allow this new generation to take part in the discussions on AI and the ethical and social issues around its development. CIÉ members were united in their response to this theme:

"AI was selected because it has cross-sectoral dimensions and encompasses issues of particular interest, since they blend science, society and the development of public policies. In this simulation, AI allowed members of the emerging generation of researchers to take part in the discussions and think about Quebec's leadership position in this area."
[translation] (participants guide, p. 5).

With this in mind, the Montréal Declaration provided three problems that had been identified during the citizen co-construction activity in the winter of 2018. We felt that it was relevant, both as part of work on the Declaration and for the work of the young researchers selected for this activity, to discuss these themes and issue recommendations. These three problems highlight particularly sensitive issues in AI development that urgently need to be debated:

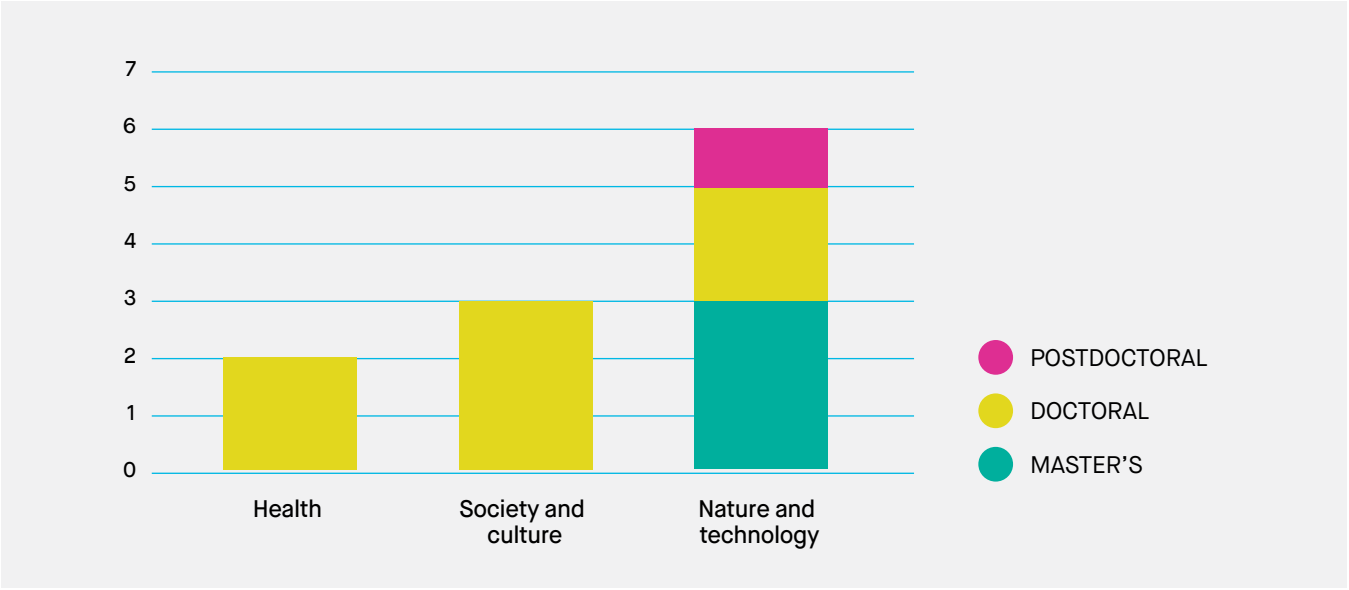
1. The security and integrity of AIS, *i.e. How can we maximize the positive impacts while minimizing the adverse effects of AI development?*
2. AI, the media and the manipulation of information, *i.e. How can we fight the dissemination and amplification of fake news and disinformation campaigns? How can we foster a democratization of access to information while encouraging critical thought and informed decision making?*
3. Public, private and participative governance: the digital commons. *Which of these types of governance is the most appropriate? What checks and balances are needed?*

The simulation had three objectives: "(1) to familiarize the participants with writing and presenting a policy brief, (2) to facilitate the acquisition of skills related to developing science policies, and (3) to analyze a social problem from a scientific point of view." [translation] (participants guide, p. 6). Policy briefs were developed in an exercise that was designed, first and foremost, to serve pedagogical purposes; the objective was not to disseminate the results to decision-makers and stakeholders. However, the recommendations in these briefs reveal the views of students from the emerging generation of researchers, and they are particularly relevant. The briefs have therefore been

attached to this report, even though they do not contain actual recommendations on public policies. We decided to present the briefs in their original form, unrevised, so as not to misrepresent their work, and in an effort to present their contribution as authentically as possible (see Appendix 2).

The activity was carried out on October 18 and 19, 2018 at the Université de Sherbrooke and involved 11 students with varying levels of education and different types of expertise.

Chart 5: Profile of Participating Students, Based on Area of Study (according to the three FRQ funding areas)



The students were assigned to three groups, with each group addressing one of the problems and led by a facilitator (someone who was independent of the Declaration, to avoid influencing the recommendations in any way). The students were given four presentations on AI and on writing policy briefs. Then they were given only six hours to draft their briefs and prepare an oral presentation on their work. The three groups presented their briefs to a jury[5] on the morning of October 19. The contest was won by Team 2 (which had been given the problem "AI, the media and the manipulation of information").

[5] The jury was consisted of three members:

Claude Asselin, Full Professor, Department of Anatomy and Cellular Biology, Faculty of Medicine and Health Sciences, Université de Sherbrooke [representing ACFAS].

Benoit Sévigny, Director of Communications and Knowledge Mobilization [representing FRQ]

Nathalie Voarino, Doctoral candidate in Bioethics, Scientific Coordinator at the Montréal Declaration [representing the Declaration]

4.2

PROBLEMS IDENTIFIED ON THE BASIS OF CITIZEN CONCERNS

Problem 1. Public Security and System Integrity

During the consultations, the citizens acknowledged that the development of AI could help make our physical and digital environments safer. For example, in an intelligent city, intelligent transportation systems may reduce traffic accident rates; in public health, epidemiological models may allow authorities to better predict the spread of illnesses; and in cybersecurity, IT security specialists are using AI to recognize attacks.

However, the citizens also recognized that certain conditions are necessary in order to ensure that AI advances are beneficial to public security. Guaranteeing "proper" use of AI through system integrity and security is fundamental to the responsible development of these technologies.

AI's negative impacts on public security can take four different forms:

1. **An AIS designed to threaten public security.**[6] For example, the use of AI for cybercrime (identity theft, hacking into nuclear power stations, etc.), political destabilization (targeted propaganda, the creation of fake videos, etc.) and the automation of military equipment (drones, robot soldiers, etc.).[7]
2. **An AIS that uses information for purposes other than those originally intended.** In this case, the citizens fear that complete medical records could be misused by insurance companies, that school records could be used to automate the labour market, or that automated traffic systems could be used to follow and monitor road users.
3. **Willful hijacking of AIS.** A person with malicious intent could directly target how the algorithm works[8], such as by outwitting a facial recognition system to gain access to protected data. Someone could also take advantage of the security challenges created by the proliferation of connected objects[9], such as to take control of an autonomous vehicle, or to paralyze a network with a massive denial-of-service attack.[10]
4. **An AIS that has been poorly evaluated:** An AIS whose reliability or robustness has been overestimated and which has caused an accident.[11] For example, the citizens mentioned that an accident involving an autonomous truck or a systematic error by a medical diagnosis program can have serious consequences.

The citizens therefore wondered how to limit the negative impacts of AI on (public) security. They raised various potential dilemmas related to system security and integrity:

> Could respect for transparency (a frequently mentioned imperative) jeopardize security by facilitating hacking?

> Does providing the most security possible necessarily mean that the system will be less

[6] Brundage M, Avin S, Clark J, Toner H, Eckersley P, Garfinkel B, et al. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. 2018; (February 2018). Available at: http://arxiv.org/abs/1802.07228

[7] Bouvet M, Chiva E. *Un regard (décalé ?) sur Intelligence Artificielle et Défense/Sécurité* - CGE [Internet]. Conférence des Grandes Écoles. 2016 [visited on Sept. 3, 2018]. Available at: http://www.cge.asso.fr/liste-actualites/un-regard-decale-sur-intelligence-artificielle-et-defensesecurite/

[8] Kurakin A, Goodfellow I, Bengio S, Dong Y, Liao F, Liang M, et al. *Adversarial Attacks and Defences Competition* [Internet]. 2018 [visited on Sept. 3, 2018]. Available at: https://arxiv.org/pdf/1804.00097.pdf

[9] Zhang Z-K, Cho MCY, Wang C-W, Hsu C-W, Chen C-K, Shieh S. *IoT Security: Ongoing Challenges and Research Opportunities.* In: 2014 IEEE 7th International Conference on Service-Oriented Computing and Applications [Internet]. IEEE; 2014 [visited on Sept. 3, 2018]. p. 230–4. Available at: http://ieeexplore.ieee.org/document/6978614/

[10] Franceschi-Bicchierai Lorenzo. *How 1.5 Million Connected Cameras Were Hijacked to Make an Unprecedented Botnet* [Internet]. Vice Motherboard. 2016 [visited on Sept. 3, 2018]. Available at: https://motherboard.vice.com/en_us/article/8q8dab/15-million-connected-cameras-ddos-botnet-brian-krebs

[11] Amodei D, Olah C, Brain G, Steinhardt J, Christiano P, Schulman J, et al. *Concrete Problems in AI Safety* [Internet]. [visited on Sept. 3, 2018]. Available at: http://arxiv.org/abs/1606.06565.pdf

efficient (it must be secure without becoming inoperative)?

> And, more generally, how can the positive impacts of AI development be maximized while preventing the adverse effects?

Other relevant references:
Asilomar AI principles
Adversarial ML

Problem 2. AI, the Media and the Manipulation of Information

The citizens were concerned about the risk that users may be manipulated, to the extent that their actions are increasingly affected by the AI mechanisms influencing their decision-making, often without their knowledge or through incentives. This raises a problem of trust in these applications, since there is a form of interference with one's autonomy, and a risk that the systems will give direction to actions (for example, based on private interests). For example, the citizens wondered whether new technologies derived from AI could create a new lobbying class, which could at times become too powerful. To maintain a certain level of freedom in the choices suggested by the AI and to avoid placing blind trust in these applications, it would therefore be important for all citizens and professionals interacting with the AI application to cultivate critical thinking skills.

Although propaganda is not a new phenomenon, it can now be created and disseminated through fake news and disinformation campaigns with unprecedented ease and speed. This includes through platforms for creating and disseminating content online (through social networks, blogs and Internet sites, and discussion forums) that is structured according to attention retention, advertising and recommendation models.[12,13,14,15] This phenomenon is also amplified by an ability to very accurately target individuals by collecting and analyzing personal data, as we saw in the Cambridge Analytica scandal[16]. This reduces the diversity of the content seen by each individual to the sum total of whatever is closest to what he or she has already liked, shared and commented on. This leaves people mainly exposed to ideas that they agree with, such that the individual is caught in a "filter bubble,"[17] raising doubts about the likelihood that any citizen today will develop critical thinking.

Each of the major social media companies has announced a series of measures to limit the propagandist potential of their tools (see the transparency reports of Facebook, Google and Twitter[18]), but is this enough? How can we ensure that these tools, which have democratized access to information and interpersonal connections, are not used to democratize propaganda? How can we fight the dissemination and amplification of fake news and disinformation campaigns, to save democracy? How can we foster the democratization of information access while encouraging critical thinking and informed decision-making?

[12] Ingram M. *Fake news is part of a bigger problem: automated propaganda* [Internet]. Columbia Journalism Review. 2018 [visited on Sept. 3, 2018]. Available at: https://www.cjr.org/analysis/algorithm-russia-facebook.php

[13] Lewis P. *"Fiction is outperforming reality": how YouTube's algorithm distorts truth.* The Guardian [Internet]. February 2, 2018 [visited on Sept. 3, 2018]; Available at: https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth

[14] Marwick A, Lewis R. *Media Manipulation and Disinformation Online* [Internet]. Data & Society Research Institute; May 2017 [visited on Sept. 3, 2018]. Available at: https://datasociety.net/output/media-manipulation-and-disinfo-online/

[15] Tusikov N. *Regulate social media platforms before it's too late* [Internet]. The Conversation. 2017 [visited on Sept. 3 2018]. Available at: http://theconversation.com/regulate-social-media-platforms-before-its-too-late-86984

[16] Cadwalladr C, Graham-Harrison E. *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach.* The Guardian [Internet]. March 17, 2018 [visited on Sept. 3, 2018]; Available at: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[17] Pariser E. *The filter bubble: what the Internet is hiding from you.* London: Penguin Books; 2012.

[18] Preliminary Facebook report: https://transparency.facebook.com/community-standards-enforcement/
Google Transparency Report: https://transparencyreport.google.com/about
Twitter Transparency Report: https://transparency.twitter.com/fr.html

Other relevant references

Caplan R, Hanson L, Donovan J. *Dead Reckoning, Navigating Content Moderation After Fake News.* Data & Society Research Institute; February 2018 [visited on Sept. 3, 2018]. Available at: https://datasociety.net/output/dead-reckoning/

Foisy P-V. *Facebook veut s'attaquer aux fausses nouvelles au Canada* [Internet]. Radio-Canada.ca. [visited on Sept. 3, 2018]. Available at: https://ici.radio-canada.ca/nouvelle/1109432/fake-news-facebook-fausses-nouvelles-canada-verification-faits

Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. *The science of fake news.* Science. March 9, 2018; 359(6380):1094-6.

Jeangène Vilmer J-B, Escorcia A, Guillaume, M, Herrera J. *Les manipulations de l'information, un défi pour nos démocraties* [Internet]. Paris, France: CAPS and IRSEM; August 2018. Available at: https://www.defense.gouv.fr/irsem/page-d-accueil/nos-evenements/lancement-du-rapport-conjoint-caps-irsem.-les-manipulations-de-l-information

Internet sites to visit

The Computational Propaganda Project: http://comprop.oii.ox.ac.uk

Observatory on Social Media: https://truthy.indiana.edu

Conversation AI: https://conversationai.github.io

"A Citizen's Guide to Fake News" Center for Information Technology & Society, UC Santa Barbara: http://cits.ucsb.edu/fake-news

Problem 3. Public, Private or Participative Governance: Digital Commons

The citizens often raised issues about how the management of AI development will be shared by public and private institutions, as well as the related risks, such as conflicts of interest, how to protect the independence of institutional actors and public institutions, the market value of data, and privacy protection.

The risk that a private monopoly will emerge in AI development management was also mentioned several times. Some participants expressed concern that monopolies would emerge, in particular since a few companies own massive amounts of data (which are needed to make AI work). Their market power is bolstered by mergers with new, smaller service providers[19].

With respect to governance by the state, a legal framework for AI comes with its own risks and challenges[20], such as being overly focused on application capabilities at the expense of protecting human values[21], such that one might wonder if it is even possible to regulate AI. This raises doubts about the real power of the state[22].

Although discussions of governance issues often place public institutions at odds with private companies, an alternative has been proposed: participative governance. This mode of governance places citizens directly in control, and may involve carrying out a major public consultation or creating a permanent consultation forum.

In the context of participative governance, the participants proposed letting users make a major contribution to the design and management of AI tools. This participation could take the form of design thinking, using open-source equipment. Such equipment, which is accessible to all, was associated with the concept of digital commons, i.e. all the shared and co-created resources and knowledge that are available free of charge (e.g. open-source software). This is more than just a form of ownership: it is a mode of cooperative organization that guarantees horizontality (exchanges between peers) and freedom of expression[23]. This type of organization depends on letting the actors themselves choose the forms of regulation.

"Digital deployment is characterised by Internet communities. This process

[19] *Big data: Bringing competition policy to the digital era* - OECD [Internet]. [cited 2018 Sep 3]. Available from: http://www.oecd.org/competition/big-data-bringing-competition-policy-to-the-digital-era.htm

[20] Scherer MU. *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies.* Harvard Journal of Law & Technology, Vol. 29, No. 2, Spring 2016. http://dx.doi.org/10.2139/ssrn.2609777

[21] Ambrose ML. *Regulating the loop: ironies of automation law.* 2014;38.

[22] Danaher J. *Philosophical Disquisitions: Is effective regulation of AI possible? Eight potential regulatory problems* [Internet]. Philosophical Disquisitions. 2015 [cited 2018 Sep 3]. Available from: http://philosophicaldisquisitions.blogspot.com/2015/07/is-effective-regulation-of-ai-possible.html

[23] Crosnier HL. *Communs numériques et communs de la connaissance. Introduction*. tic&société. May 31, 2018;(Vol. 12, N° 1):1–12.

presupposes the emergence of significantly new organizational forms supported by information technologies, in particular open-source movements, and Web 2.0."[24] [translation]

This mode of governance is not without its challenges, including the fact that it is vulnerable to various forms of enclosure (fewer common uses), instituted by the state as well as by companies[25].

These issues raise a number of questions: What is the best way for AI governance to be shared between public, private and participative management? Does tension necessarily exist between these different modes of management, and which is the most appropriate? Is it necessary to benchmark these types of management, and if so, which types of guidelines should be put in place?

Other relevant references

Chessen M. *Encoded laws, policies, and virtues: the offspring of artificial intelligence and public-policy...* [Internet]. Medium. 2017 [cited 2018 Sep 3]. Available from: https://medium.com/artificial-intelligence-policy-laws-and-ethics/encoded-laws-policies-and-virtues-the-offspring-of-artificial-intelligence-and-public-policy-3dfb357faf9

Shafto, P. *Why Big Tech Companies Are Open-Sourcing Their AIS* [Internet]. IFLScience. [cited 2018 Sep 3]. Available from: https://www.iflscience.com/technology/why-big-tech-companies-are-open-sourcing-their-ai-systems/

Internet sites to visit

The FACIL website

The Déclaration des communs numériques de FACIL

4.3.

RECOMMENDATIONS FROM THE NEW GENERATION OF RESEARCHERS

This exercise was particularly fruitful, and the three briefs provide relevant recommendations on responsible AI.

The first policy brief examines the consequences of a poor evaluation of AI capabilities, a problem that is an integral part of the first problem proposed on system security and integrity. The purpose of this brief is to promote security and protection of Canadians and, more specifically with respect to embedded systems. The brief recommends creating a Canada-wide organization for certifying embedded systems that use AI (ARBIA) and developing a no-fault liability plan. The brief is notable for the relevance of its recommendations, which refer to existing mechanisms (MEI[26] and integration of the recommendations into Quebec's Digital Strategy), as well as for proposing an original mechanism for protecting citizens (either the state or the businesses that have developed such embedded systems will be liable for the damages arising from an accident).

The second brief, entitled "Fake news, real issues: educating ourselves to confront the issues" [translation], examines the problem of information on the Internet that has been manipulated using AI (in particular, the creation of fake news). Here the students underscored the need to encourage more critical thinking. Their recommendations are in line with the importance of educating Quebec citizens about media and information. Their analysis led to two main recommendations:

[24] Ruzé E. *La constitution et la gouvernance des biens communs numériques ancillaires dans les communautés de l'Internet. Le cas du wiki de la communauté open-source WordPress*. Management & Avenir. 2013;(65):189–205.

[25] Crosnier HL. *Une bonne nouvelle pour la théorie des biens communs*. Vacarme. 2011;(56):92–4.

[26] Formerly MESI (*Quebec's Ministère de l'économie*, de la science et de l'innovation), which was renamed MEI (*Ministère de l'économie et de l'innovation,* or the department of economics and innovation) on the day that the brief was written.

1. the Quebec public should be warned about fake news (vigilance), including from specialized AIS, and

2. Quebec teachers should be provided with the tools they need to raise awareness about fake news through the education system, starting in primary school.

In order to ensure that these mechanisms are implemented, they recommend creating a Quebec Digital Vigilance Committee, under the aegis of the International Observatory on the Societal Impacts of Artificial Intelligence and Digital Technologies, and integrating educational processes into Quebec's Digital Action Plan.

The third brief examines the problem of AI governance, exploring issues related to the gaps in current policies and regulations concerning the AI sector. Here the challenge is to find a form of AI governance that will best respond to the needs of the various actors affected (businesses, citizens and public institutions). The main objective of this policy brief was to *provide a method for working on and thinking about AI governance* in order to appropriately address the issues raised: the inadequacy of current policies, citizens' concerns, the lack of clarity when an incident occurs, and the absence of methods for managing AI-related problems. Several potential solutions were proposed. The students recommended creating an independent (provincial) organization to regulate the use of common data that will be responsible for, among other things, implementing educational mechanisms, as well as adjusting current laws and regulations to address the new technological realities. The brief also proposed integrating a component into the organization's mission so that it will be constantly in phase with the market. The organization in question, named "Educ'AI" in the oral presentation, would be set up as a think tank.

Summarizing, the students recommend implementing an independent organization to manage AI, with a variety of responsibilities (involving the creation of a certification or a system of vigilance), as well as educational processes. These recommendations are consistent with those formulated during other co-construction activities. While implementing an independent organization or educational mechanisms is aligned with the Declaration's recommendations for public policies, the recommendation to create a system of responsibility merits further analysis. Echoing the recommendations made at the citizen forums last winter on implementing insurance mechanisms that would set parameters for the sharing of responsibility when there is fault (see *Part 3: Summary report of the recommentations from the winter co-construction workshops*), this recommendation suggests a full-fledged analysis of general and criminal responsibility for the impacts of AI.

Lastly, the relevance of this activity has led us to support the CIÉ in its recommendation that more opportunities should be created for graduate students to be trained in non-academic professional activities and, more specifically, in political participation.

5. CONCLUSION

These three activities allowed us to explore issues related to AI development under new themes (e.g. propaganda), but also to experiment with a new approach (e.g. a simulation activity).

Independent of the activity, the participants' recommendations support the need to implement training that is tailored to everyone's needs, to update the legal and regulatory framework, and to develop new knowledge on AI developments and their impacts. These recommendations also encourage the promotion of participatory governance, with an emphasis on the importance of involving the stakeholders at different key moments in the management of AI development and in policy decision making.

By opening a discussion of new potential solutions and sectoral differences, these activities suggest that co-construction deserves to be pursued beyond the work carried out by the Montréal Declaration, and they support the relevance of a public consultation on responsible AI.

APPENDIX 1
The Paris Scenarios (in French only)

DÉMOCRATIE

Fausse nouvelle dans la campagne électorale

23 mars 2022. Ce matin, Dominique B. se rend à la réunion de crise de l'Agence sur l'intégrité de l'information (AII), mise en place dans le cadre de la Loi contre la manipulation de l'information. Le président de la République sortant, candidat à sa réélection, vient de perdre 7 points dans les sondages d'intention de vote en trois semaines et la tendance à la baisse semble se confirmer. Alors qu'il était assuré de l'emporter deux mois auparavant, il est désormais dépassé par la candidate populiste de droite qui a pris la tête de la course électorale. Le tournant se situe le 2 mars, avec la diffusion sur internet d'une vidéo montrant le président de la République discuter avec le président du Mouvement des entreprises de France, en marge de son école d'été. Le président de la République assurait qu'il comprenait la situation des entreprises qui employaient des travailleurs immigrés sans papiers, qu'il était important de maintenir des bas salaires pour garantir la vitalité des petites et moyennes entreprises, et qu'il veillerait à ce que ces entreprises ne soient pas pénalisées.

La vidéo s'était vite répandue dans les réseaux sociaux et les propos du Président avaient été relayés dans les premières heures par deux grands médias, la chaîne d'information TBT et le site lefureteur.com. Le porte-parole de l'Élysée avait immédiatement démenti les propos attribués au Président et avait fait savoir que la vidéo était un faux créé par une agence étrangère qui tentait d'interférer dans les élections françaises. La technique utilisée pour créer la vidéo avait été mise au point par l'entreprise américaine Monkeypaw Productions qui avait tiré parti des algorithmes GAN (*generative adversarial networks*), élaborés par des chercheurs de l'Université de Montréal en 2014. Contre toute attente, les images créées grâce à l'IA avaient atteint un degré de réalisme stupéfiant en moins de dix ans, si bien qu'une fausse vidéo ne pouvait plus être détectée à l'œil nu.

Ni le démenti de l'Élysée, ni le mea culpa de TBT et du Fureteur, ni encore l'interdiction de diffusion de la vidéo n'avaient eu l'effet espéré. La vidéo était encore consultable sur différents sites étrangers comme le site rassvet.io. Un député du parti populiste de droite en avait profité pour accuser le Président de faire le jeu de l'immigration clandestine et de nuire aux intérêts des Français. Le nombre de gazouillis avec le mot-clic *#Presidentclandestin* avait passé la barre des 300 000 en une semaine. À un mois du premier tour des élections présidentielles, Dominique B., directrice de l'AII, doit présenter un plan pour enrayer les effets dévastateurs de cette fausse information et rétablir les conditions d'une campagne électorale saine. Mais ce matin, le sentiment d'avoir déjà épuisé toutes les solutions l'emporte à l'AII.

ENVIRONNEMENT

La cote environnementale basée sur l'empreinte de carbone

1er février 2025. Pour la cinquième année de suite les températures battent des records de chaleur dans le monde entier. La majorité des pays ayant signé l'Accord de Paris en décembre 2015 n'ont pas tenu leurs engagements en raison des impératifs économiques de court terme, malgré les mises en garde du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC). En conséquence, les villes européennes du C40, le réseau des villes engagées dans la transition écologique, ont accéléré leur coopération pour proposer à leurs habitants un système de permis carbone individuel fortement incitatif, le système ÉcoFit, connecté à leur compte bancaire et aux différentes applications d'achat en ligne : dans ces villes, le prix des biens et services est affiché en euros et en carbone, et chaque citoyen doit viser 4 tonnes d'émission de carbone par an pour l'ensemble de sa consommation. Les personnes qui atteignent cet objectif augmentent leur cote environnementale calculée par l'algorithme ÉcoFit, à partir de leurs données personnelles de consommation. Cette cote leur donne un accès gratuit à de multiples services écoresponsables en transport, éducation, formation et culture.

15 juin 2025. Au moment de passer leur commande de coquilles Saint-Jacques grâce à leur réfrigérateur FrigoMax connecté, Ive et Charles, habitants du 20e arr. à Paris, découvrent ce nouveau système de points auquel ils viennent d'adhérer : Coquilles Saint-Jacques (Provenance : Pérou) : 12 € / 22 kg éq. CO2/kg*[27]. Un message d'avertissement s'affiche : « Cet achat doit rester exceptionnel. Vous ne pourrez pas tenir votre objectif annuel si vous le reproduisez souvent. » Et l'algorithme de recommandation de FrigoMax leur propose alors des coquilles Saint-Jacques de Saint-Brieuc, fraîches, qui coûtent 22,5€ mais seulement 0,25 kg éq. CO2/kg.

15 octobre 2025. Après quelques écarts, et suite aux nombreux messages d'avertissement, Ive et Charles ont fait un effort pour consommer plus sobrement grâce aux recommandations d'ÉcoFit : régime presque végétarien, nouvelle isolation de leur logement, transport en commun et en vélo, contrat d'électricité verte, choix exclusif d'applications avec *data-centers* carbone neutre : c'est qu'au bureau, tout le monde compare maintenant sa cote environnementale !

1er décembre 2025. Grâce à leurs comportements de plus en plus vertueux, Ive et Charles ont réussi à rester juste en-dessous du plafond visé : après 6 mois, ils sont chacun à 1,95 tonne de carbone pour leur consommation globale. De plus, ils ont moins dépensé monétairement, ce qui leur procure une épargne inattendue. Le couple considère alors de réaliser son projet de séjour à Cuba pour Noël et commence à consulter les sites des agences de voyage. Un message leur parvient sur leur téléphone : « Attention à l'*effet rebond* : dépenser vos économies dans un voyage annulerait tous vos efforts ! Pensez à voyager local ! »

[27] kg éq. CO2/kg = kilogramme équivalent carbone ; exprimé ici par kg de produit importé par avion.

MONDE DU TRAVAIL

Forage des données (*data mining*) RH pour optimiser l'ambiance au travail

30 octobre 2025. Pierre-André a enfin décroché un emploi dans un bon bureau d'avocats qui traite notamment du droit de l'environnement, l'un de ses domaines de prédilection.

Après trois semaines de travail, il rencontre Marco aux ressources humaines pour une séance de mentorat personnalisée. Marco fait le point sur l'intégration de Pierre-André, sur ses attentes initiales, ses difficultés, etc. Il lui explique aussi que la firme utilise désormais AmbIA+, une IA d'analyse conversationnelle qui étudie les attitudes des salariés et aide à maintenir une ambiance de travail apaisante et productive. C'est une question d'efficacité. Ainsi, tous les courriels, appels téléphoniques et prises de parole en réunion d'équipe sont analysés pour extraire un historique des humeurs et des émotions des salariés. Ces données sont ensuite rapportées à un laboratoire de recherche en psychologie.

Pierre-André est déstabilisé et même un peu inquiet, mais Marco essaie de le rassurer :

> AmbIA+ fournit une assistance individualisée, elle conseille et entraîne, mais il n'y a pas de sanction. D'ailleurs, AmbIA+ ne mémorise que la forme des interactions, et tous les échanges que vous avez eus jusqu'à présent au bureau se sont bien passés.

Tous, sauf pour le 15 et le 16 octobre derniers. Pierre-André travaillait alors sur le dossier de la nouvelle station d'épuration des eaux usées de la ville de Lille. « Selon AmbIA+, rapporte Marco, vous avez à plusieurs reprises interrompu vos collègues en réunion pour répéter les mêmes idées, ce qui a créé de la tension chez eux. Il faudrait essayer d'exposer vos arguments en une fois, lors du tour de table, pour ne pas perdre de temps. »

Mais ce n'est pas tout :

> Apparemment, l'algorithme a aussi détecté des périodes d'inactivité sur le réseau de plusieurs heures, sans aucun échange avec vos collègues. Ce n'est pas grave en soi, mais c'est mieux de maintenir le contact avec l'équipe. Est-ce que vous vous souvenez de la raison de cette inactivité ?

Pierre-André n'est plus seulement inquiet, il est embarrassé et s'interroge sur la pertinence de ces questions :

> Oui, c'est vrai, j'aime bien travailler avec un crayon sur un rapport papier et je préfère ne pas rédiger directement sur le document collaboratif en ligne… et en effet lors de la réunion du 15, j'apportais une idée nouvelle qui ne me semblait pas bien comprise et je craignais qu'on ne l'oublie. Mais est-ce vraiment un problème ?

Compréhensif, Marco répond qu'il n'y a vraiment aucun problème : « Mais ne vous déconnectez pas de l'équipe, c'est mieux pour la performance collective. Allez, on se revoit dans deux mois. Et bonne chance pour la réunion de demain ! »

APPENDIX 2
Student policy briefs

Simulation 2018, CIÉ-FRQ

Brève politique sur l'intelligence artificielle

Le présent document est le résultat d'un exercice de simulation, dont l'objectif était d'acquérir des compétences en rédaction et en communication publique. Étant donné le contexte pédagogique dans lequel cette note a été produite, elle n'a pas la vocation, dans les faits, d'être adressée à des décideurs ou à des acteurs de la fonction publique.

La Déclaration de Montréal a choisi de publier ces brèves afin de représenter fidèlement le résultat d'un travail réalisé en 6 heures par les étudiants de la relève et montrer la pertinence d'un tel exercice.

Problématique 1 : Sécurité publique et intégrité des systèmes

Sous-problématique 4 : Les conséquences engendrées par une mauvaise évaluation des capacités de l'intelligence artificielle

Document rédigé par :

Joël Simoneau

Jérôme Gélinas Bélanger

Fidele Ndjoulou

Moumouni Ouiminga

À l'intention du Gouvernement du Canada

Titre de la brève

Pour l'établissement d'un système de responsabilité de l'intelligence artificielle dans les biens de consommation

Cette brève politique expose une démarche qui vise à promouvoir une intelligence artificielle responsable pour la sécurité et la protection des Canadiennes et Canadiens. Elle porte spécifiquement sur les systèmes embarqués. Les recommandations énoncées sont :

1. La mise sur pied d'un organisme pancanadien de certification des systèmes embarqués[1] utilisant l'intelligence artificielle
2. Le développement d'un régime de responsabilité sans faute

De manière concrète, le gouvernement devrait s'atteler dans un premier temps à la création d'un organisme fédéral responsable de la certification obligatoire des systèmes embarqués utilisant l'IA. Dans un deuxième temps, il est impératif de mettre sur place un régime propriétaire sans faute.

Une telle politique permettra d'assurer une meilleure santé et sécurité ainsi qu'une protection légale à tous les Canadiennes et les Canadiens dans leur interaction avec des objets utilisant l'IA, tout en impliquant les entreprises privées dans le processus de la saine utilisation de l'IA.

[1] On qualifie de « système embarqué » un système électronique et informatique autonome dédié à une tâche précise, souvent en temps réel, possédant une taille limitée et ayant une consommation énergétique restreinte. www.futura-sciences.com/tech/definitions/technologie-systeme-embarque-15282/

La société canadienne à l'ère du développement de l'intelligence artificielle

Notre société est en train de vivre une transformation globale basée sur l'évolution du numérique. Autant celui-ci véhicule des informations à une vitesse précédemment inimaginable, qu'il transforme notre rapport avec les objets. Les avancées technologiques récentes en intelligence artificielle (IA) permettent d'imaginer un futur imminent où certaines tâches avec prises de décision redondantes seraient attribuées à des logiciels conçus expressément pour cette fonction. Les véhicules autonomes sont déjà au coin de la rue, les dispositifs médicaux intelligents sont derrière les portes des universités. Une mauvaise médicamentation ou des accidents automobiles sont des dangers qui tendent à être réglés par l'utilisation intelligente et sécuritaire de l'IA, mais il faut aussi s'assurer qu'elle n'en devient pas la cause. Cela représente des inquiétudes énoncées par les Canadiennes et les Canadiens à travers les travaux de la Déclaration de Montréal pour un développement responsable de l'IA.

La régularisation des systèmes embarqués, soit un appareil physique contenant un logiciel utilisant une IA, devrait être un projet d'importance pour le gouvernement canadien. Ceux-ci représentent une implémentation physique et commercialisable d'un produit d'IA, et il serait important d'en assurer une règlementation en amont de leur arrivée prochaine sur le marché canadien. Une prise de décision proactive et l'installation d'un cadre réglementaire permettrait l'encadrement des IA pouvant avoir un impact physique direct sur le peuple canadien.

Ce document propose l'instauration d'un organisme réglementaire de certification des systèmes embarqués utilisant l'IA et d'un régime de responsabilité basé sur le propriétaire sans faute. L'organisme permettrait d'encadrer les normes de sécurité de conception et d'utilisation des systèmes, et le régime permettrait de définir exactement le rapport de responsabilité dans le but de protéger les Canadiennes et les Canadiens, autant légalement qu'au niveau de leur santé et bien-être. La combinaison de ces deux mesures encadrera les systèmes embarqués, de leur commercialisation jusqu'à leur utilisation, ce qui maximisera les impacts positifs du développement de l'IA, en réduisant ses effets néfastes.

Constats et pistes d'action sur le développement de l'intelligence artificielle au Canada

1. Organisme de certification

À l'heure présente, aucun cadre législatif n'existe quant à l'utilisation de l'intelligence artificielle intégrée à des systèmes embarqués au Canada. Ce flou juridique pose un certain nombre de défis pour les différents paliers de gouvernement, notamment le gouvernement fédéral, relativement à leur capacité de structurer la mise en marché et la régulation de ces objets au pays. De façon plus générale, ce manque de structure à ce niveau engendre des complexités juridiques en termes d'évaluation du risque que présentent ces technologies pour le public, mais aussi en termes de l'attribution du poids de la responsabilité advenant un incident découlant de l'utilisation d'une technologie basée sur l'IA.

Face à ces défis, il apparaît nécessaire pour l'État canadien de créer un organisme réglementaire de certification des systèmes embarqués utilisant l'IA, l'office de réglementation nommé l'Agence de Réglementation sur les Biens utilisant l'Intelligence Artificielle (ARBIA), à vocation interdisciplinaire et agissant comme pilier décisionnel. Cet organisme possédera trois principaux axes d'action afin de parvenir à structurer la réglementation de l'IA à l'échelle canadienne: 1) l'investissement dans la recherche et l'innovation permettant le développement de balises législatives basées sur des connaissances techniques, 2) l'instauration de comités experts possédant une bi-spécialisation reposant sur l'IA et leur propre champ d'expertise à l'intérieur des différents ministères pouvant être éventuellement affectés par le développement de l'IA et 3) le développement d'une plateforme réglementaire encadrant la mise en marché et le régime propriétaire sans faute.

L'implication directe du gouvernement canadien dans les cas de problématique de bien utilisant l'IA permettra d'assurer une veille scientifique et sécuritaire proactive, et de protéger légalement les consommateurs canadiens, qui n'auront pas à subir des procès-bâillons.

2. Régime de responsabilité sans faute

On entend par responsabilité l'obligation de répondre d'un dommage devant la justice et d'en assumer les conséquences notamment civiles et pénales envers la victime et/ou la société. Dans un régime de responsabilité sans faute, le gouvernement du Canada sera responsable des accidents physiques ou matériels causés par un bien matériel utilisant l'IA. Dans le cas d'un bien non conforme au processus de certification, le gouvernement canadien peut intenter des actions contre le fabricant.

L'implication directe du gouvernement canadien dans les cas de problématique de bien utilisant l'IA permettra d'assurer une veille scientifique et sécuritaire proactive, et de protéger légalement les consommateurs canadiens, qui n'auront pas à subir des procès-bâillons.

2.1 Secteurs d'activités concernés

L'intelligence artificielle s'applique à plusieurs secteurs d'activité notamment la santé, l'éducation, la sécurité, l'agriculture. Cependant, cette brève politique touche de manière spécifique l'automobile autonome, les dispositifs médicaux et la domotique.

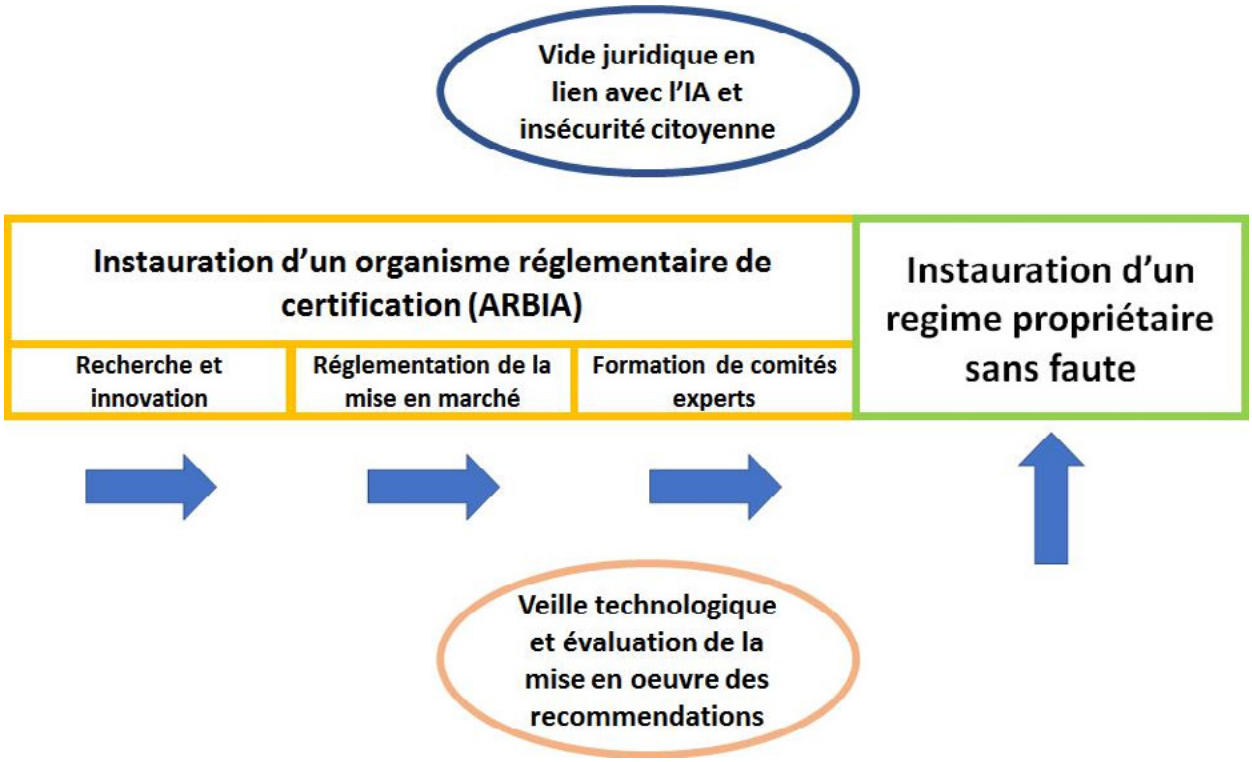
2.1.1 Automobiles autonomes

Avec le développement de l'intelligence artificielle, le secteur de l'automobile a connu une transformation radicale. Une nouvelle catégorie d'automobile dite automobile autonome est mise sur le marché. Néanmoins, ces voitures ont déjà causé des accidents aux États-Unis, par exemple l'accident mortel causé par une voiture autonome en Floride en mars 2018. Compte tenu de l'absence des règles spécifiques à la circulation des automobiles autonomes, et dans le souci d'apporter une meilleure protection aux citoyens, il est impératif de définir un cadre réglementaire au niveau fédéral. Ce cadre va fixer la responsabilité des parties prenantes, à savoir l'État et les compagnies propriétaires des voitures autonomes. Une faute liée aux défaillances est une faute de la compagnie responsable et propriétaire de la voiture, tandis qu'une utilisation faite par l'individu utilisateur est une faute de sa part.

2.1.2 Dispositifs médicaux

Pour faciliter la vie des personnes vivant avec le diabète, une pompe à insuline a été développée. C'est un système intégré composé de trois parties (boîtier,

composantes électroniques, cathéter) qui libère automatiquement de l'insuline. Le fonctionnement de ce système nécessite impérativement une formation du patient, une autosurveillance glycémique et un suivi médical rapproché. Cela engendrera une responsabilité du patient en cas de mauvaise utilisation de sa part. Les conséquences pourraient être énormes au point d'engendrer d'éventuels cas de décès.



Retombés et recommandations

Dans l'objectif d'assurer la santé et la sécurité des Canadiennes et Canadiens face aux systèmes embarqués utilisant l'IA et de maximiser les impacts positifs du développement de l'IA, nous émettons les recommandations suivantes.

Recommandation 1

Création de l'Agence de Réglementation sur les Biens utilisant l'Intelligence Artificielle (ARBIA)

Retombées :

- Coût nul pour le gouvernement canadien

  Le financement de l'organisme de certification et des actions légales sera couvert par une licence de fabrication des systèmes.
- Fiabilité des systèmes d'IA pour la santé et sécurité des Canadiennes et Canadiens.

  Par le respect de normes définies par des comités experts, normes qui seront mises à jour selon les cas vécus.

Recommandation 2

Instauration d'un régime propriétaire sans faute

Retombées :

- Accessibilité judiciaire améliorée dans le cas de faute des fabricants.

  Le système sans faute donne la responsabilité de la poursuite judiciaire au gouvernement canadien, qui a plus de ressources que les citoyennes et citoyens individuellement.
- Promotion de l'implication sociale des entreprises.

  Considérant leur responsabilité directement impliquée, les fabricants vont être encouragés à développer des mécanismes d'utilisation sécuritaire de leurs produits.
- Veille technologique et sécuritaire du gouvernement canadien

  Considérant l'implication directe du gouvernement canadien dans les processus judiciaires, celui-ci assure une veille permanente dans la gestion saine des IA.

Bibliographie :

[1] Pour aller plus loin voir Palmer, Vernon. « Trois principes de la responsabilité sans faute » (1987) 39:4 Revue internationale de droit comparé; Mémeteau, Gérard. « Un point sur la responsabilité civile du fait des prothèses » (2013) 2013:123 Médecine & Droit 175-180; Jacob, Julien. « Prévention des risques technologiques à l'aide de la responsabilité civile en présence d'une innovation à double impact » (2013) 202:1 Économie & amp; prévision 1-18.
[2] https://diabetnutrition.ch/les-traitements/la-pompe-a-insuline-quest-ce-que-cest/
3https://ici.radio-canada.ca/info/videos/media-7560667/premier-accident-mortel-impliquant-une-voiture-autonome source consultée le 18 octobre 2018

Simulation 2018 dans le cadre des J2R

« Politique et intelligence artificielle »

Le présent document est le résultat d'un exercice de simulation, dont l'objectif était d'acquérir des compétences en rédaction et en communication publique. Étant donné le contexte pédagogique dans lequel elle a été produite cette note, elle n'a pas la vocation, dans les faits, d'être adressée à des décideurs ou à des acteurs de la fonction publique.

La Déclaration de Montréal a choisi de publier ces brèves afin de représenter fidèlement le résultat d'un travail réalisé en 6 heures par les étudiants de la relève et montrer la pertinence d'un tel exercice.

Fausses nouvelles, vrais enjeux : s'éduquer pour y faire face

Présenté aux membres du jury

Par

Jean Clairemond César, étudiant au doctorat en éducation, Université de Sherbrooke
Isabelle Dufour, inf., candidate au doctorat, Université de Sherbrooke
Gaël Grissonnanche, post-doctorant en physique, Université de Sherbrooke
Philippe Lebel, doctorant en microbiologie, Université de Montréal

18 octobre 2018

Tables des matières

But de la brève

Émettre des recommandations auprès d'un décideur public en vue d'offrir une ou plusieurs pistes de solution à un problème spécifique découlant d'une des trois problématiques décrites dans le document élaboré par l'équipe de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle. Une problématique sera attribuée par équipe et les participants devront identifier les éléments suivants :

- Le **problème**
- La ou les **solution**(s) recommandée(s) et les **répercussions** sur la population visée et non visée
- Le **décideur public** impliqué
- Les **facteurs environnementaux** pouvant faire obstacle à la mise en œuvre de la ou des solution(s) recommandée(s).

Couverture (1 page)

La première page présente une synthèse de la brève politique. Elle présente la pertinence de la brève et ses grandes lignes, les conclusions clés et la marche à suivre.

Cette brève politique présente le problème des fausses nouvelles sur l'internet. Aujourd'hui la proportion de Canadiens qui consomme de l'information en ligne a dépassé celle des médias traditionnels. L'efficience de cette technologie repose sur l'intelligence artificielle (IA) en offrant des contenus filtrés selon le comportement et l'intérêt de l'utilisateur. De nos jours, plusieurs acteurs sociopolitiques ont levé le drapeau rouge sur cet enjeu de société. D'un autre côté, les grandes entreprises de médias sociaux telles que Google, Facebook, Amazon et tant d'autres proposent déjà des mesures pour limiter le potentiel propagandiste de leur algorithme, et la définition d'une fausse nouvelle ne fait pas consensus. La pertinence de notre brève se situe dans la nécessité d'augmenter l'esprit critique au sein de la population québécoise. L'absence d'esprit critique peut occasionner plusieurs problèmes en éducation et en santé et dans d'autres domaines. Destiné aux ministres concernés par la stratégie numérique, ce document présente plusieurs recommandations sur l'importance de l'éducation aux médias et à l'information au Québec.

Introduction (1 page)

Cette section décrit l'objectif principal de la brève et le problème politique. Elle établit un lien entre les données probantes et le problème.

L'avènement de l'internet apporte aux citoyens la démocratisation de l'accès à l'information à travers des moteurs de recherches intelligents et des médias sociaux. Aujourd'hui, la proportion de Canadiens consommant de l'information en ligne a dépassé celle des médias traditionnels. L'efficience de cette technologie repose sur l'intelligence artificielle (IA) en offrant des contenus filtrés selon le comportement et l'intérêt de l'utilisateur. Tandis que certaines études montrent que cette exposition partiale à l'information tend à engendrer chez l'utilisateur une confirmation systématique de sa pensée, d'autres en revanche arguent que celui-ci n'a jamais été exposé à une telle diversité de sources lorsque comparé à la presse écrite, à la télévision, à la radio, etc.

C'est dans ce contexte qu'émerge sur la scène internationale la notion de fausse nouvelle comme un enjeu de désinformation massive dans une société démocratique. L'usage d'IA comme en a fait la firme Cambridge Analytica aux États-Unis a montré au monde le niveau de déstabilisation sociétale que cette technologie peut engendrer. Alors que les grandes entreprises de médias sociaux telles que Google, Facebook, Amazon et tant d'autres proposent déjà aujourd'hui des mesures pour limiter le potentiel propagandiste de leur algorithme, la définition d'une fausse nouvelle ne fait pas consensus. En effet, selon le Global News, près de 58% des Canadiens définissent celle-ci comme une histoire pour laquelle les faits sont faux. Cependant, 46% l'emploient pour désigner les nouvelles de journaux et les discours de personnalités politiques n'exprimant qu'un unique côté des faits. Encore, ce même chiffre désigne le pourcentage pensant que ce terme est uniquement utilisé par les politiciens pour discréditer les médias qui les critiquent. À l'autre bout du spectre, des actions pour valoriser l'esprit critique de l'utilisateur demeurent une avenue qui doit être envisagée.

La problématique amenée par l'essor des fausses nouvelles dans les médias est importante et est susceptible d'avoir un impact important sur la population. À cet égard, l'objectif de cette brève est d'augmenter la sécurité de la population et leur éducation face aux fausses nouvelles.

Données probantes et analyse (3 pages)

Cette section représente le cœur de la brève politique. La qualité de cette section est jugée par la pertinence des données présentées, des interprétations tirées de ces données, ainsi que de leurs apports et de leurs limites. Elle peut contenir des graphiques, des tableaux et des schémas.

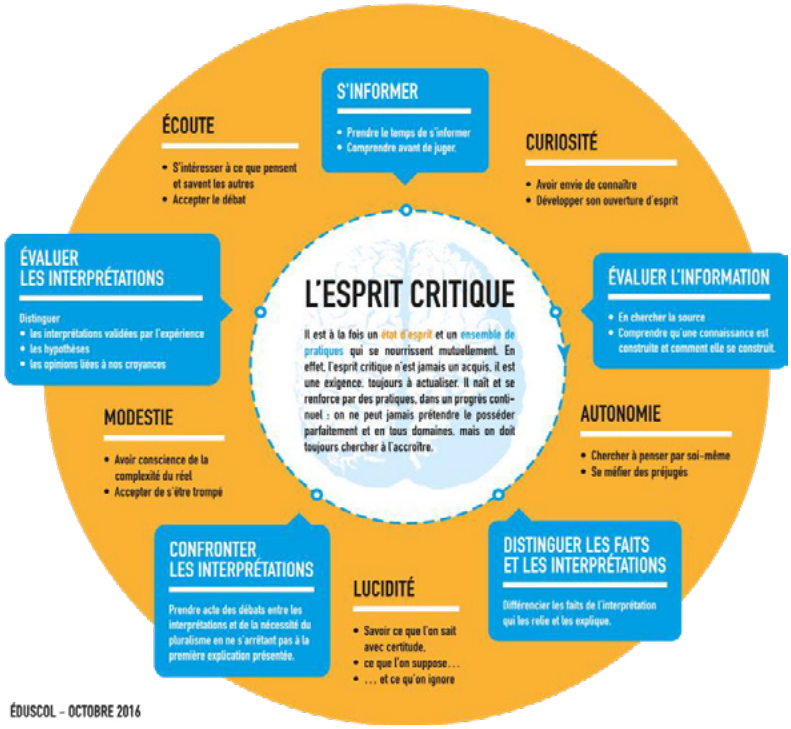
Selon Jeff Yates, expert québécois de la question, une fausse nouvelle se définit comme « une information soit carrément fausse, détournée, exagérée ou dénaturée à un point tel qu'elle n'est plus véridique, et présentée comme une vraie nouvelle dans le but de tromper les gens. Cela peut être fait pour générer des clics et des partages sur les réseaux sociaux, pour atteindre des objectifs quelconques (politiques, idéologiques, économiques, etc.) ou simplement pour se moquer de la crédulité des lecteurs ». Sujet de débats socio-économiques, les fausses nouvelles dans les médias ont connu un essor marqué durant les dernières années, et principalement avec le développement de l'IA. En effet, des méthodes associées à l'IA sont utilisées par les sites de médias sociaux et peuvent procéder de façon automatique à la diffusion de fausses nouvelles.

En Amérique du Nord, environ 60% de la population croit que la dispersion de telles informations dans les médias cause de la confusion. Les enfants et les adolescents sont particulièrement à risque d'attribuer du crédit et de participer à leur diffusion. Le manque de consensus dans la définition d'une fausse nouvelle contribue à l'incertitude vécue par la population, 45% des Canadiens en ayant une vision erronée. À l'ère numérique, la nécessité de développer une pensée critique concernant les informations transmises par les médias se positionne donc comme un enjeu central.

Selon Vallerand, la pensée critique est une pensée responsable qui s'appuie sur des critères et qui est sensible au contexte et aux autres (Vallerand, 2016). L'IA pourrait être utilisée pour développer l'esprit critique des jeunes et les former au doute constructif. Les jeunes du Québec apprendront comment mettre en perspective une information diffusée sur le web. En ce sens, les décideurs politiques donneront les moyens nécessaires pour y arriver.

En France par exemple, le développement de l'esprit critique est au centre de la mission assignée au système éducatif français, comme le présente le modèle de l'esprit critique d'Eduscol. Il est renforcé par l'attention désormais portée à l'éducation aux médias et à l'information. Le travail de formation des élèves au décryptage du réel et à la construction, progressive, d'un esprit éclairé, autonome et critique est essentiel. L'esprit critique est une compétence essentielle du citoyen et de la citoyenne du 21^{e} siècle. Analyser une source, mettre en perspective une image ou une information, en extraire l'essentiel, critiquer le contenu, se questionner sont autant de savoirs numériques nécessaires à l'exercice d'une citoyenneté avisée. Une étude faite en 2017 au Royaume-Uni a montré que seulement 4 % de la population testée avait été capable d'identifier correctement les vraies des fausses nouvelles. Ce résultat est inquiétant, notamment en termes de sécurité publique. Pensons par exemple au mouvement anti-vaccination qui

cause un retour en force de maladies mortelles, tel que la coqueluche aux États-Unis, malgré qu'il a été démontré depuis longtemps que les causes de ce mouvement sont fausses.



Selon les experts de la pensée critique, Christopher DiCarlo et l'auteur du *Petit cours d'autodéfense intellectuel*, Normand Baillargeon, la solution passe par l'éducation. En effet, il est préférable que l'école forme une jeunesse plus critique de ce qu'elle consulte plutôt que de faire confiance aux grandes entreprises privées du web pour autocensurer leur contenu.

D'ailleurs on retrouve plusieurs initiatives, ailleurs comme chez nous, qui vont dans ce sens. En France, plus précisément en Haute-Savoie, des enseignantes ont créé une habitude locale où elles prennent une heure par semaine pour sensibiliser leurs élèves à détecter les fausses nouvelles. Cette initiative, accueillie avec enthousiasme par les élèves, semble rapidement porter fruit puisque ces jeunes de 10 ans ont déjà développé les réflexes de vérifier d'où proviennent des images-chocs qui publicisent de fausses nouvelles sensationnalistes, par exemple.

Plus près de chez nous, depuis mai 2018, un nouveau programme s'implante dans les écoles ontariennes : Actufuté. Ce programme se veut une collaboration entre la Fondation pour le journalisme canadien et l'organisme CIVIX qui est responsable du programme Vote étudiant. Ce dernier prend vie autour des périodes d'élections et encourage la participation citoyenne des 9 à 19 ans. C'est dans ces périodes riches en nouvelles qu'Actufuté viendra aider les élèves à démystifier le vrai du faux.
Au Québec, « École branchée », un organisme sans but lucratif (OSBL) propose des outils aux enseignantes et aux enseignants pour intégrer ces considérations dans leur programme de tous les jours. Malheureusement, à ce jour, seulement 15 % à 20 % du corps enseignant est rejoint par l'organisme. Démonstration qu'une intervention gouvernementale est nécessaire pour offrir une protection équitable à tous nos jeunes contre ce fléau. Ce faisant, la jeunesse pourra aussi transmettre cette information et conscientiser ses proches à la problématique.

Pour y arriver, le Plan d'action numérique en éducation et en enseignement supérieur, annoncé à l'été 2018, prévoit quelque 900 millions de dollars pour, justement, préparer la génération de demain à ce nouvel environnement numérique. Le gouvernement du Québec pourrait ainsi soutenir les services d'« École branchée », voire même intégrer son contenu au cursus normal de l'éducation primaire et secondaire.

L'intégration de cette notion d'éducation directement au cursus scolaire vient contrer l'obstacle environnemental principal. Le désir des professeurs d'assurer le développement de leurs étudiants pourra aussi agir à titre de facilitateurs.

Répercussions sur les politiques et recommandations (1 page)

Cette section présente les recommandations proposées et les répercussions anticipées. Ces recommandations et ces répercussions peuvent s'organiser autour de thèmes, de parties intéressées ou d'échéancier.

L'argumentaire soulevé met en lumière plusieurs défis soulevés par l'IA. Entre autres, elle contribue à la diffusion de masse de fausses nouvelles. Cela cause de la confusion au sein de la population, une perte de confiance envers les sources d'information. Les jeunes et les adolescents sont particulièrement sensibles aux fausses nouvelles, leur capacité de raisonnement et leur esprit critique étant en construction. L'implication des instances gouvernementales est donc primordiale pour assurer la protection et l'éducation des populations, et particulièrement des jeunes, sur la problématique des fausses nouvelles. Les recommandations adressées font appel à des notions de vigilance et d'éducation.

Vigilance

Notre recommandation :

- Alerter la population québécoise sur la dissémination de fausses nouvelles

L'objectif étant de favoriser le développement et l'utilisation d'IA spécialisée pour détecter les fausses nouvelles diffusées sur les médias sociaux.
Cette initiative s'inscrit également en parallèle avec le Détecteur de rumeurs, où les alertes du CVMQ, en cas de détection d'une fausse nouvelle de grande importance, pourront être diffusées.
La création du Comité de vigilance numérique du Québec sera annexée à l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique.

Éducation

Notre recommandation :

- Offrir des outils au corps enseignant québécois pour intégrer la conscientisation face aux fausses nouvelles dans l'éducation, dès l'école primaire.

Cette mesure aura deux buts : préparer directement cette génération à affronter le fléau des fausses nouvelles et les inciter à répandre ces bonnes pratiques auprès de leurs proches.

Grâce au financement déjà prévu pour le Plan d'action numérique en éducation et en enseignement supérieur ainsi qu'aux initiatives déjà en place, il sera possible de protéger la population québécoise sans investissement supplémentaire et sans réinventer la roue. À court terme, la promotion de ces outils auprès des enseignantes et des enseignants aura déjà un impact et il sera possible de penser intégrer ces enseignements au cursus normal à moyen terme.

Tableau 1. Grille d'évaluation des brèves politiques

Critères	Tous les points	- 1 p	- 2 p	- 3 p
Couverture	La synthèse présente la pertinence de la brève et ses grandes lignes, les conclusions clefs et la marche à suivre.	Des éléments sont manquants	La synthèse est manquante	
Introduction	Cette section décrit très bien l'objectif principal de la brève et le problème politique. Elle établit un lien entre les données probantes et le problème.	Cette section décrit bien l'objectif principal de la brève et le problème politique. Elle établit un lien entre les données probantes et le problème.	Cette section décrit convenablement l'objectif principal de la brève et le problème politique. Elle établit un lien entre les données probantes et le problème.	Cette section est absente
Données probantes et analyse	Cette section est très pertinente au regard du problème; Les interprétations sont justes et convaincantes; Les facteurs environnementaux (socio-politico-économico-culturels) sont très bien pris en compte dans la possible intégration des recommandations; Les apports et les limites sont très bien identifiés.	Cette section est pertinente au regard du problème; Les interprétations sont justes; Les facteurs environnementaux (socio-politico-économico-culturels) sont bien pris en compte dans la possible intégration des recommandations; Les apports et les limites sont bien identifiés.	Cette section est plutôt pertinente au regard du problème; Les interprétations sont plutôt justes; Les facteurs environnementaux (socio-politico-économico-culturels) sont pris en compte dans la possible intégration des recommandations; Les apports et les limites sont identifiés.	Cette section n'est pas pertinente au regard du problème; Les interprétations sont erronées; Les facteurs environnementaux (socio-politico-économico-culturels) ne sont pas pris en compte dans la possible intégration des recommandations; Les apports et les limites ne sont pas correctement identifiés.
Répercussions et recommandations	Les recommandations proposées sont très pertinentes et les répercussions anticipées très bien identifiées.	Les recommandations proposées sont pertinentes et les répercussions anticipées sont bien identifiées.	Les recommandations proposées sont plus ou moins pertinentes et les répercussions anticipées plus ou moins bien identifiées.	Les recommandations proposées ne sont pas pertinentes et les répercussions anticipées ne sont pas bien identifiées.
Qualité de la présentation orale	La présentation de la brève est très convaincante.	La présentation de la brève est convaincante.	La présentation de la brève est peu convaincante.	La présentation de la brève n'est pas convaincante.
Total des points	/15			

Simulation 2018 dans le cadre des J2R

« Politique et intelligence artificielle »

Le présent document est le résultat d'un exercice de simulation, dont l'objectif était d'acquérir des compétences en rédaction et en communication publique. Étant donné le contexte pédagogique dans lequel a été produite cette note, elle n'a pas la vocation, dans les faits, d'être adressée à des décideurs ou à des acteurs de la fonction publique.

La Déclaration de Montréal a choisi de publier ces brèves afin de représenter fidèlement le résultat d'un travail réalisé en 6 heures par les étudiants de la relève et montrer la pertinence d'un tel exercice.

Gouvernance publique, privée ou participative : les communs numériques

Présenté aux membres du jury

Par

Thomas Bousquet
Alexandre Côté, PhD(c)
Christian Kouakou, PhD(c)

18 octobre 2018

Tables des matières

Couverture

Vue d'ensemble

L'évolution rapide de la technologie et la science entourant l'intelligence artificielle (IA) exposent certaines brèches dans les politiques et les règlementations actuelles qui concernent ce secteur de développement. Le gouvernement doit se pencher sur ce problème qui soulève de vives inquiétudes pour la population qui s'interroge sur la protection de sa vie privée. L'inquiétude reste présente du côté des entreprises privées qui elles, ont incessamment besoin d'alimenter leur système d'IA avec des données de plus en plus complexes et précises. Le défi du gouvernement est de trouver une forme de gouvernance de l'IA qui répond au mieux aux besoins des différents acteurs concernés.

L'objectif principal de cette brève politique proposée par notre Groupe de travail ponctuel sur l'utilisation de l'IA et des données communes est donc de *fournir une méthode de travail et de réflexion* afin de répondre adéquatement à ces interrogations, en s'assurant de considérer les intérêts distincts de la population et du secteur privé.

Intérêts des acteurs impliqués

Population	- Protéger ses données personnelles - Être rassuré par l'indépendance des instances faisant usage de ses données
Gouvernement	- Assurer la protection du public - Stimuler la croissance économique du secteur technologique
Privé	- Connaître une croissance économique stable - Développer et améliorer les connaissances concernant l'IA

Problèmes soulevés par ces intérêts distincts :

- Politique actuelle mal adaptée
- Inquiétudes des citoyens
- Flou dans les responsabilités lors d'incidents
- Absence de méthodes pour gérer les problèmes liés à l'IA

Plusieurs pistes de solutions

- Création d'un organisme de régulation indépendant
- Favorisation d'une responsabilité partagée relativement aux données communes
- Ajustement des lois et réglementations en vigueur afin de les adapter aux nouvelles réalités technologiques

Introduction

L'évolution rapide de la technologie et la science entourant l'intelligence artificielle (IA) exposent certaines brèches dans les politiques et les règlementations actuelles qui concernent ce secteur de développement. Bien que certaines lois soient déjà en place, la vitesse de l'appareil public peut difficilement rattraper celle de la croissance technologique, et les règles en place deviennent rapidement inadaptées.

Les inquiétudes de la population

Cette inadéquation des politiques publiques en matière d'encadrement de l'IA et de l'utilisation des données servant à sa croissance inquiète la population québécoise. Une consultation récente initiée par un groupe d'experts composé, entre autres, de gens de l'Université de Montréal, de l'Université McGill et de l'Institut de valorisation des données (IVADO), a permis d'identifier certaines préoccupations clés des citoyens vis-à-vis les enjeux actuels concernant notamment :

- La responsabilité face aux données et à l'IA ;
- La protection de la vie privée des individus ;
- La valeur marchande des données partagées ;
- Les risques de mise en place d'un monopole, et de conflits d'intérêts entre les différents acteurs touchés ;
- Ainsi que l'indépendance des différents acteurs qui interviennent dans le domaine.

Les considérations face au secteur privé

La croissance économique québécoise étant de plus en plus liée aux nouvelles technologies, à l'exploitation des données et au développement de l'IA, il est important pour le gouvernement – malgré les inquiétudes soulevées – de ne pas laisser le secteur privé au dépourvu. L'accès aux données de la population est le carburant de cet important moteur économique qui doit manifestement être régulé, mais pour qui une marge de manœuvre doit être maintenue.

Le défi de la gouvernance

Les trois acteurs généraux qui sont touchés par la problématique – le gouvernement, la population et le secteur privé – ressentent déjà les impacts du manque d'ajustement des politiques actuelles. L'exemple récent des piratages de données des grands services technologiques comme Facebook et Google – utilisés par des centaines de milliers de Québécois – expose bien ce phénomène : la population est ultimement la victime, blâmant à la fois le secteur privé et le gouvernement pour les dommages encourus.

Dans cette optique, il est donc primordial pour le député délégué à la transformation numérique gouvernementale de se positionner et de répondre aux questions et inquiétudes soulevées par la société civile et le secteur privé :

1. Qui est responsable face aux données communes ?
2. Quel appareil assure la protection du public ? Fonctionne-t-il adéquatement ?
3. Quel niveau de transparence est optimal ?
4. Comment peut-on stimuler la croissance économique dans le secteur des nouvelles technologies, tout en maintenant la confiance de la population face au processus ?

L'objectif principal de cette brève politique proposée par notre Groupe de travail ponctuel sur l'utilisation de l'IA et des données communes est donc de *fournir une méthode de travail et de réflexion* afin de répondre adéquatement à ces interrogations, en s'assurant de considérer les intérêts distincts de la population et du secteur privé.

Données probantes et analyse

Le problème de l'immigration discriminante

Le gouvernement canadien serait en expérimentation de l'utilisation de l'IA pour le tri des demandes de visa et d'immigration. Cette information a été publiée dans un rapport du Citizen Lab et relayée dans les médias canadiens.

L'une des auteurs du rapport a souligné que « sans garanties et mécanismes de surveillance appropriés, utiliser l'IA pour déterminer l'immigration et le statut de réfugié est très risqué ». Il est clair que l'utilisation de l'IA pourrait aider grandement à accélérer le tri et traitement de données d'immigration. Cela ne saurait cependant et en aucun cas outrepasser la décision discrétionnaire liée au droit à l'immigration qui ne saurait être laissé à une machine ou un algorithme ; algorithme qui plus est, n'est pas sans biais car fortement dépendant des considérations des personnes qui programment. Un arbitrage est à faire entre « rapidité dans le traitement des demandes » et « sélection discrétionnaire des dossiers ».

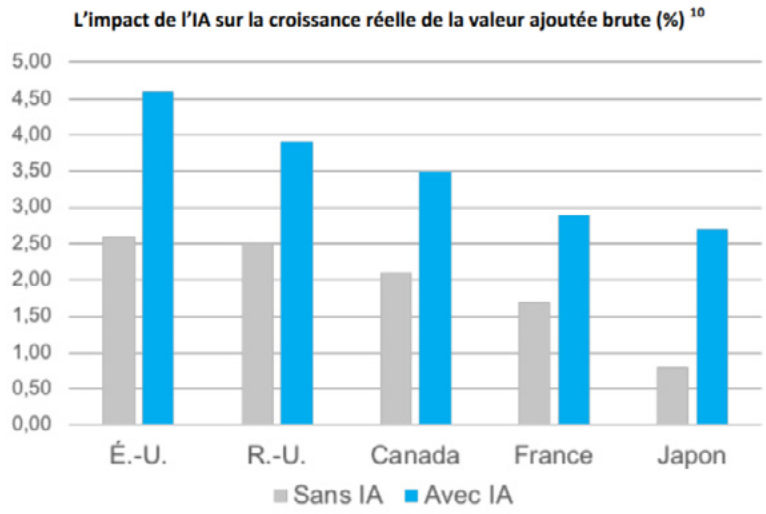
En outre, le rapport à l'immigration de l'équipe responsable de l'algorithme pourrait fortement déteindre sur le résultat final de sélection. D'une part, l'on pourrait avoir un processus de sélection moins rigoureux, voire laxiste, qui laisserait entrer sur le territoire québécois des personnes ne remplissant pas les conditions requises pour l'immigration. D'autre part, un processus plus strict pourrait ôter la possibilité aux personnes remplissant les conditions d'y accéder. Car rappelons-le, la sélection initiale aura été faite non pas par choix discrétionnaire, mais par une machine intelligente.

La confidentialité des données

Les plus gros acteurs en matière de la gestion des données d'utilisateurs, dont font partie Google et Facebook, mettent en place des actions correctives dans leur gestion des données personnelles : ces données ont une valeur monétaire importante et ne doivent être utilisées que dans le cadre où elles ont été fournies par l'utilisateur du service. Les piratages sont récurrents comme l'ont montré les évènements récents (500 000 comptes Google et 29 millions d'utilisateurs Facebook piratés) et la population est donc inquiète de l'utilisation qui peut être faite de ses données.

Les *Big Data* ont également un intérêt dans la santé des populations : des intelligences artificielles sont en développement pour aider au diagnostic médical et nécessitent donc des données sensibles, très personnelles et donc qui font partie d'un haut niveau de confidentialité. La société québécoise a donc évoqué ses inquiétudes quant à la gestion de ces données lors de l'étude qui a été menée sur une partie de la population lors de la rédaction de la Déclaration de Montréal pour l'IA responsable.

Les données personnelles sont donc classifiées selon leur confidentialité et sont donc disponibles à plusieurs niveaux. L'utilisateur sait à qui et dans quel but chaque organisme collecte des données à son sujet.
Il devient donc important qu'un comité multisectoriel neutre se charge de veiller à ce que les possesseurs de ces grandes quantités de données les utilisent de manière éthique pour éviter que des centaines de milliers d'utilisateurs québécois de ces services voient leurs données diffusées sur le web.



De nombreux pays font du développement de l'IA une priorité majeure en investissant dans des organismes de recherche visant à progresser dans ce domaine, comme le montre ce graphique issu de la communication de « Économie, science et innovation Québec » à propos de l'essor de l'écosystème québécois en intelligence artificielle publié en mai 2018.

Le Québec ressent également ce besoin et prévoit des fonds à la recherche dans ce domaine, c'est pourquoi cette volonté doit alors s'étendre à la protection des populations et de leurs données sans pour autant empêcher les acteurs privés et universitaires de progresser dans leurs recherches et leurs innovations.
D'après le projet de mettre le numérique au service du bien commun au Québec (disponible sur le site économie.gouv.qc.ca) d'ici 5 ans, le Québec prévoit mettre à la disposition de la population une transformation numérique des municipalités qui se traduit par une collecte de données continue.

De plus, le Québec prévoit également que les citoyens pourront interagir de façon numérique avec les services de santé et sociaux d'ici les prochaines années. Toutes les données nécessaires à ces services nécessitent des données sensibles sur les citoyens et il est donc très important pour la province de se mobiliser pour protéger les utilisateurs contre une mauvaise utilisation de ces données dans tous les domaines confondus, que ce soit les services publics ou bien les entreprises privées de services.

Le développement du numérique, prisé par le Québec, doit donc faire évoluer les règlements relatifs à la protection des données des citoyens en faisant travailler les entreprises leaders de l'IA conjointement avec les pouvoirs publics et les intérêts des utilisateurs.

Répercussions sur les politiques et recommandations

Nous constatons au final un manque d'adaptation des politiques actuelles face à l'IA et la gestion des données qui alimentent sa croissance. Cette problématique engendre un flou qui touche non seulement le gouvernement et la classe politique, mais également la population et le secteur privé.

Il existe actuellement un manque de clarté quant à la responsabilité de ces trois acteurs face aux données communes (1). Non seulement la *Loi sur la protection des renseignements personnels et des documents électroniques* (LPRPDE ; fédérale) et la *Loi sur la protection des renseignements personnels dans le secteur privé* (provinciale) ne répondent pas à ce manque, elles ne se sont pas non plus adaptées assez rapidement aux nouvelles réalités technologiques, ce qui risque de créer un doute dans la population quant à sa protection et la protection de ses données personnelles (2). Un certain degré de transparence (3) est évidemment requis afin de pallier cet aspect de la problématique, tout en gardant à l'esprit l'importance de ne pas entraver le développement économique du secteur technologique au Québec.

Recommandations

Dans cette optique :

1. Nous emboîtons le pas du *Citizen Lab* de l'Université de Toronto et recommandons la création d'un organisme indépendant de régulation de l'utilisation des données communes. À la différence des chercheurs torontois, nous recommandons cependant que cet organisme soit de juridiction provinciale afin de prendre en considération les particularités de la population québécoise.

2. Nous favorisons une responsabilité partagée des données communes, alimentée par ce nouvel organisme. Ce dernier serait chargé, entre autres, de l'éducation de la population en matière de protection des données personnelles et de la surveillance du secteur privé quant à l'utilisation de ces données.

3. La création de cet organisme viendrait également répondre à la deuxième question soulevée par notre analyse de la situation, soit l'identité de l'appareil veillant à la protection du public. Il serait maintenant clair aux yeux de la population qu'une entité veille à ses intérêts, entre autres en s'assurant – à travers des recommandations émises à l'endroit du gouvernement – de l'adéquation des lois et réglementations concernant l'IA et la gestion des données.

4. Nous suggérons fortement que le nouvel organisme s'assure qu'un certain niveau de transparence soit respecté, autant par le gouvernement que par le secteur privé. Le type de données utilisées, leurs sources, les buts de leur utilisation et leur portée devraient être de nature publique.

5. Nous recommandons qu'un volet économique soit intégré à la mission de l'organisme afin que celui-ci soit constamment en phase avec le marché, s'assurant que les politiques mises en place permettent d'atteindre le parfait équilibre entre croissance et respect du milieu.

6. Finalement, nous recommandons que le nouvel organisme développe une méthode de réflexion permettant d'adapter les lois et réglementations concernant l'IA et les données communes aux changements rapides et fréquents inhérents au secteur des hautes technologies.

FINAL REPORT CREDITS

The Montréal Responsible AI Declaration was prepared under the direction of:

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

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

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

Coordination

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

Content contribution

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

Revising and editing

Chantal Berthiaume, Content Manager and Editor

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

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

Translation

Rachel Anne Normand and François Girard, Linguistic Services

Rebecca Sellers, Copywriter, Translator, ESL Teacher

Graphic design

Stéphanie Hauschild, Art Director

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

OUR PARTNERS

Université de Montréal

CRE
CENTRE DE RECHERCHE EN ETHIQUE

LAB VILLE PROSPECTIVE

IVADO

SAT

Mila

CIFAR
AI &
Society
Program

Université de Montréal | designNsociété

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

Canada

Centre Culturel Canadien Paris
Canadian Cultural Centre Paris

MEC
Maison des étudiants canadiens

< >

Montréal Declaration Responsible AI_

</ >

montrealdeclaration-responsibleai.com