< >
Montréal Declaration
Responsible AI_
</ >

PART 3

SUMMARY REPORT OF RECOMMENDATIONS FROM THE WINTER CO-CONSTRUCTION WORKSHOPS

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.

TABLE AND FIGURES

WRITTEN BY

NATHALIE VOARINO, Scientific Coordinator, PhD Candidate in Bioethics of Université de Montréal

CAMILLE VÉZY, Doctoral student in communication, Université de Montréal

VALENTINE CROSSET, Doctoral student in criminology, Université de Montréal

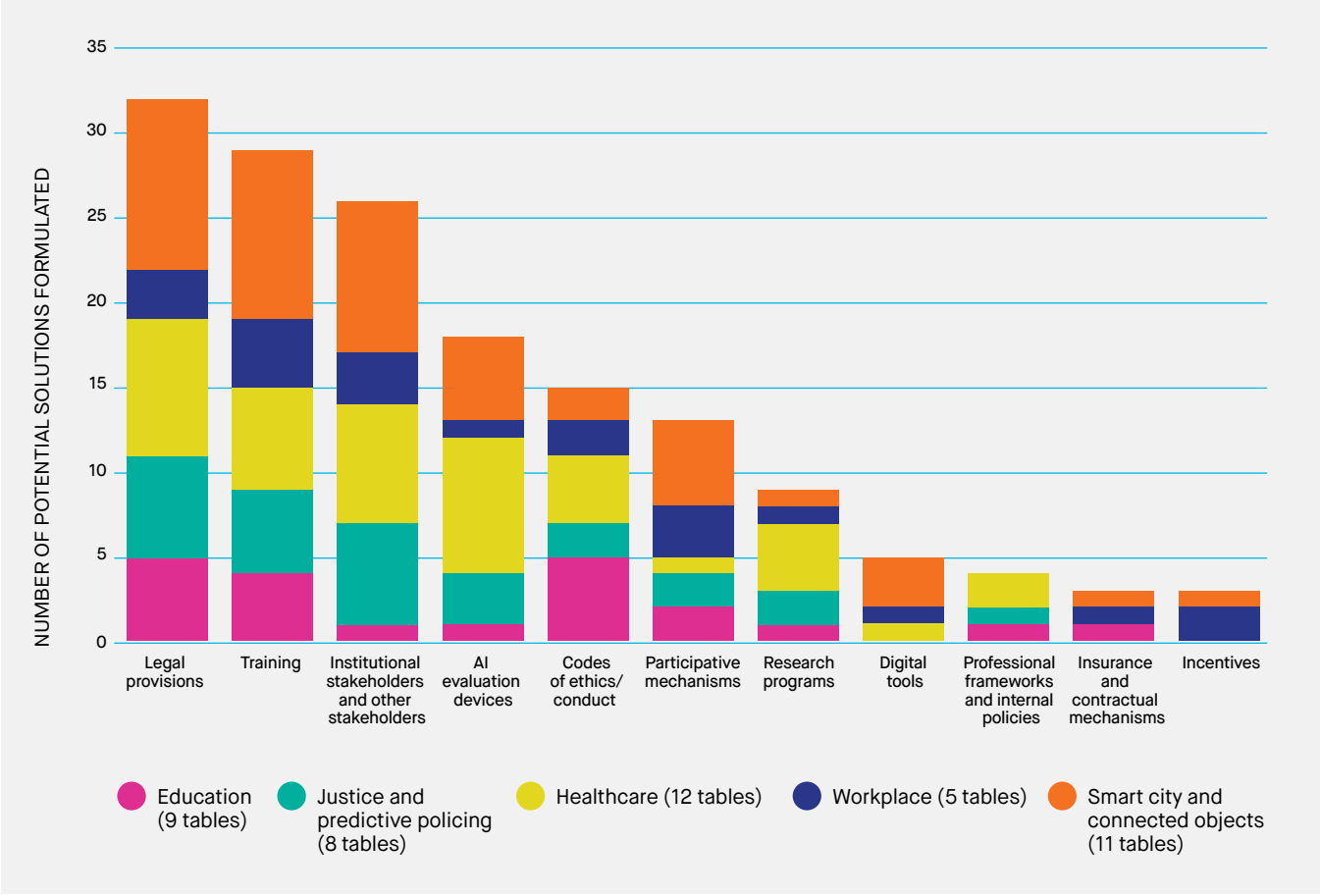
ALESSIA ZARZANI, Ph.D in Planning, Université de Montréal and Ph.D in Landscape and Environment, Université la Sapienza de Roma

1. SUMMARY

Citizens met at 45 tables to discuss their perception of risks and issues in the responsible development of artificial intelligence (AI).

They formulated 11 categories of potential solutions to these risks and issues:

Table 1: Potential solutions proposed to respond to the issues identified



2. CO-CONSTRUCTION DATA: EXPLANATORY REMARKS

The current section explains the results collected during the co-construction tables held in winter 2018 for the Montréal Declaration. There were 45 discussion tables in total that brought together hundreds of citizens. Discussions were held around five major sectors of AI development: education (9 tables); justice and predictive policing (8 tables); healthcare (12 tables), workplace (5 tables), and the smart city and connected objects (11 tables). The analysis presented in this section was also enhanced by discussions in different satellite activities (input during classes; world cafés addressing the same themes without strictly following the method used during the co-construction tables).

To understand this section properly, we should note that the discussions addressed issues not only related to responsible AI development, but also those pertaining to data management (especially personal data and big data)—whether data from which algorithms learn, or data that, in some shape or form, is analyzed by AI. As these issues are interrelated, they were considered together for this analysis.

The scenarios served as launchpads for discussions during which two types of data were collected: perceptions of citizens regarding the **risks and issues** in AI development and **potential solutions** to address them (see scenarios, Part 1, Section 6, Annex 2). For the purposes of this section, the analysis remains descriptive and as verbatim as possible. The **main directions** expected in terms of responsible AI development refer to citizen recommendations that are not specified in concrete potential solutions. They nonetheless allow us to identify the main positions and standard expectations citizens have of AI development. When these expectations were debated during discussions or when citizens considered that responding to these expectations was an issue, they were considered in the issues category.

Each co-construction table was invited to choose two or three issues to be treated as priorities before 2025. Only issues that citizens considered priorities were taken into account in Section 3 for the purposes of this report. These priority issues were defined by citizens and classified, for each sector, according to the principles of the first version of the Declaration to which they refer. However, it is worth noting that just because certain issues were not considered priorities that they were not discussed, that they are less important, or that the principles

were not discussed for each sector. One single principle for each sector is detailed in this section.

A thematic analysis was made of all the discussions using NVivo software. The purpose of this analysis was to highlight citizens' perceptions of the scope of the risks and issues in AI development (see mind map, Table 3). These issues have been grouped into 12 categories, and are not mutually exclusive. We recognize that this is but one of many ways in which to classify the different discussions that took place. The potential solutions identified by citizens to address these issues were classifed into 11 main categories. These categories are mutually exclusive, thus allowing us to add quantitative data.

With regard to the quantitative data in this report, the number of times it occurs corresponds to the number of tables where each issue/potential solution was formulated through consensus, in keeping with the co-construction process. The total number of potential solutions (n=190) corresponds to those identified as priorities by citizens (since they were invited to clearly formulate them on posters). However, potential solutions that were mentioned during the discussions but did not explicitly appear on the posters were also taken into consideration.

Quotes from the report are presented so that they reference the co-construction table when they formulated by a group (consensus). Other quotes correspond to individual ideas formulated (written on Post-its by participants or copied verbatim by team members).

3. MAIN DIRECTIONS EXPECTED BY CITIZENS

Generally speaking, the participants recognized that AI had important potential benefits. Participants recognized the time savings that AI devices could bring particularly when it came to work and legal matters:

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

However, it was also mentioned that AI had to be developed with caution and from now on, to prevent harmful use although some consider the possibilities that AI opens up to still be limited. Introducing a framework was therefore recognized as necessary to prevent risks rather than trying to determine who is to blame when they occur:

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

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

Most participants recognized the need to align public and private interests to prevent monopolies from emerging, or limit the influence of corporations (which are sometimes seen as ungovernable) through more cohesive and legal measures. To the greatest extent possible, these mechanisms should be simple and evolve so they can adapt to the pace of AI development and maintain steady control of it. In the legal sector, certain participants mentioned a "divide" between technology (defined as quick, innovative, even abstract) and our institutions (often too rigid in their integration of technology) that are not able to deal with these changes in society. Some tables went as far as suggesting "nationalizing AI", which would then "become a public service, and programmers would be public servants". (Smart city and connected objects table, INM, Montréal, February 18, 2018, Connected refrigerator scenario.)

The participants also recommended ensuring that AI be considered in context, meaning different parameters must taken into account (e.g. mandatory or optional collection of data the algorithm learns from). These mechanisms should come from and involve independent, trained people to promote diversity and include those who are the most vulnerable, and protect different lifestyles.

Whatever the use, most participants insisted that AI must remain a tool, and that the final decision be made by a human being (whether a legal ruling, hiring decision or health diagnosis), which implies recognizing its limitations.

> "AI proposes, mankind disposes."
> — A participant

Protecting an individual's privacy and managing personal data were discussed in depth. For example, processing healthcare data should be managed in a unique way, given the highly sensitive nature of the information. It should therefore both promote control methods ranked according to type of use and adopt security as an operational mode. As for the workplace sector, participants recommended that employers be obligated to inform users of how their data is processed.

The participants were aware that these recommendations involve important institutional changes, and underlined the fact AI is not necessarily desirable to begin with.

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

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

4. CITIZEN PERCEPTION OF RESPONSIBLE AI DEVELOPMENT ISSUES

4.1. INTRODUCTION

Citizens that took part in the co-construction days were invited to select two or three issues to address as priorities before the year 2025 with regard to responsible development of artificial intelligence.

Table 2: Priorities identified by citizens according to the principles of the Declaration (number of tables).

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

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

The principles of knowledge, responsibility, privacy, justice and democracy are presented below per sector. The autonomy principle, often selected as a priority, concerns preserving, even encouraging individual autonomy when faced with risks of technological determinism and reliance on tools.

It also raises the issue of the two sides to freedom of choice: being able to make your own choice when faced with a decision guided by AI, but also being able to choose not to use these tools without risking social exclusion. The freedom included in this autonomy principle regarding AI systems would involve any person's capacity for self-determination.

"Develop technologies that promote human autonomy and freedom of choice."
(Education table, Bibliothèque de Laval, March 24, 2018, Hyper-personalization of education scenario).

The well-being principle also holds an important place for participants. Participants at every table expressed a collective desire to move towards a society that is fair, equitable and promotes everyone's development. Well-being is therefore both a collective (touching on equity and accessibility issues within the justice principle) and an individual issue, aiming for everyone's fulfillment without hampering autonomy and privacy. Participants showed a preference for AI development "that would allow any individual to achieve personal and social fulfillment". (Education table, Bibliothèque Père Ambroise, Montréal, March 3, 2018, AlterEgo scenario.)

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

MAIN ISSUES DISCUSSED PER SECTOR

EDUCATION

Six out of nine tables considered privacy, responsibility, well-being and knowledge issues priorities for the education sector. Discussions on issues related to the knowledge principle were especially relevant to broaching the subject of transforming human skill sets in the age of AI:

ISSUES RELATED TO THE KNOWLEDGE PRINCIPLE
(6 out of 9 tables)

For the theme of education, issues related to the knowledge principle concern changes in skill sets, given that the teaching profession and ways of developing and accessing knowledge are rapidly changing. This principle was mostly discussed from the perspective of how the learning relationship would change, how teachers' expertise would be challenged and how their work would have to change as a result. It was also mentioned in relation to the diversity principle: the need to cultivate a wide range of intelligences and relationships to knowledge.

"Redefining/transforming the nature of the relationship between teachers and students in the classroom and changing our relationship to knowledge."
(SAT Table, Montréal, March 13, 2018, Nao scenario).

"Human skills and abilities: the importance of developing many learning environments."
(Musée de la civilisation table, Québec City, April 6, 2018, AlterEgo scenario).

LEGAL SYSTEM AND PREDICTIVE POLICING

Five out of eight tables considered privacy, responsibility and knowledge issues priorities for the justice and predictive policing sector. Discussions on issues related to the responsibility principle allowed us to clarify the principle's scope:

ISSUES RELATED TO THE RESPONSIBILITY PRINCIPLE
(5 out of 8 tables)

The responsibility principle was formulated in two primary ways: as a demand for human accountability in legal rulings, and a concern for responsibility in decision-making (and any potential errors). From the citizens' point of view, the algorithm's lack of transparency goes against accountability, since it is difficult to know what factored into the decision. The responsibility principle is therefore linked to knowledge and transparency principles in that decisions should be explainable and preserve the skills and role of human beings in the legal system.

"[Justice] must remain a tool whose sole purpose is to protect individuals. Promote compassionate and equitable justice that accounts for idiosyncrasies and past experiences. Artificial intelligence must not have the right to judge human behaviour. The final decision must always require human intervention."
(SAT table, Montréal, March 13, 2018, Preventive arrest scenario).

"Transparency, accountability and responsibility when creating the tool, the data used, and the impact of this tool."
(SAT Table, Montréal, March 13, 2018, Parole scenario).

With regard to responsibility, citizens were concerned about overlooking human beings and human "agency". Failing to consider human dynamics and the ability for invidividuals to change shows a clear concern about a "static" vision of human beings provided by an algorithm, which would make its decisions problematic and unreliable. Participants were ready to make "agency" a principle of the Declaration in this workshop.

"We must take personal agency into consideration. The ability of each individual to change, to change their own course."

HEALTHCARE

Privacy and responsibility principles were considered priorities by 9 and 10 tables out of 12, respectively, in healthcare. Privacy issues were particularly significant for the sector given the relatively sensitive and invariably personal nature of healthcare data.

PRIVACY PRINCIPLE ISSUES
(9 out of 12 tables)

Participants identified different issues related to confidentiality and invasion of privacy. At issue was the possible invasion of privacy linked to developing and configuring AI systems (e.g. which should help avoid pirating, shortages and harmful use). Citizens also discussed "retroactivity" (use of data previously collected for another purpose) and accessing this data through private companies. In light of these issues, citizens' concerns included how to ensure that data isn't sold, and how to guarantee that the patient maintains control of their data (especially when it concerns private data), and holds full rights to it.

"To what extent are we willing to share our personal data (information) as individuals in order to feed healthcare services?"
(Musée de la civilisation table, Québec City, April 6, 2018, Digital twins scenario).

WORKPLACE

Issues on justice and knowledge were considered priorities for the workplace sector, (5 and 4 tables out 5, respectively). All tables that discussed AI development in the workplace, therefore, felt that issues concerning justice, equity and diversity should be addressed separately.

ISSUES RELATED TO THE JUSTICE PRINCIPLE
(5 tables out of 5)

Citizens had two primary concerns about the justice principle: ensuring an equitable sharing of AI benefits among all social groups and territories, and "including nondiscriminatory algorithms that favour diversity, inclusion and social justice". (Musée de la civilisation table, Québec City, April 6, 2018, AI as a compulsory step to employment scenario).

"Sharing AI benefits (productivity gains); equity among social groups, territories (cities and regions), taking vulnerabilities into consideration; the meaning of work in society and how it shapes our identities."
(Musée de la civilisation table, Québec City, April 6, 2018, Socially responsible restructuring scenario).

SMART CITY AND CONNECTED OBJECTS

For the smart city and connected objects sector, issues related to autonomy and democracy principles were considered priorities by 9 and 7 tables out of 11, respectively. Citizens felt that many issues could impact the democracy principle:

ISSUES RELATED TO THE DEMOCRACY PRINCIPLE
(7 out of 11 tables)

Participants discussed issues such as balancing collective interests and individual needs; managing access to public spaces and sharing said spaces, or even sharing the benefits from the development of AI technologies (particularly between individuals, the public sector and the private sector). They insisted on a need for and the difficulty of ensuring a collective (involving citizens) and enlightened (which implies a level of transparency in developing AI systems) decision-making process to define guidelines on connected objects. Citizens also questioned the true independence of public authorities in AI development, and discussed the risk of normalizing behaviour that could lead to marginalization, thereby possibly jeopardizing the democracy principle.

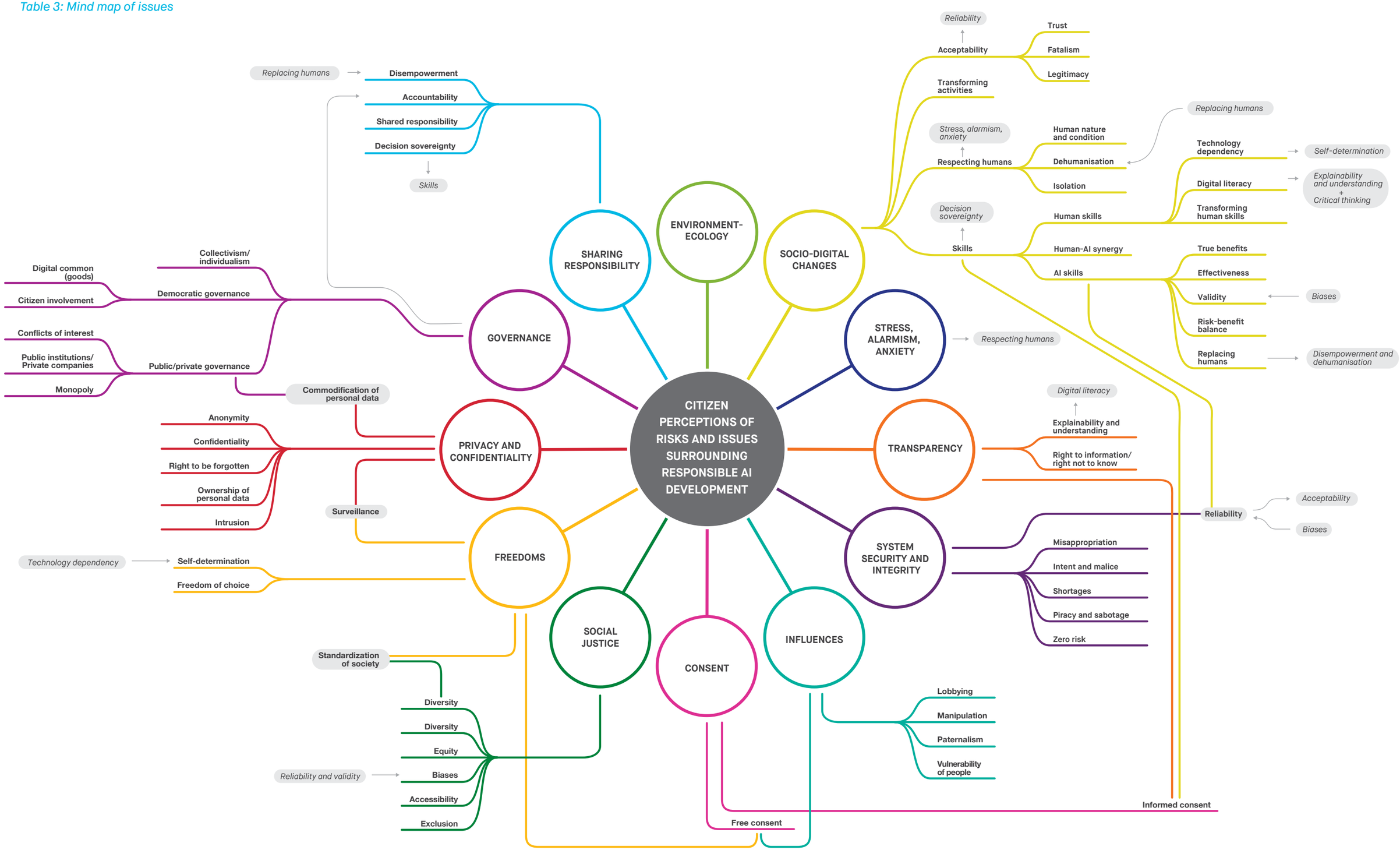
"How can we manage an intelligent transportation system democratically?"
(Bibliothèque du Boisé table, Montréal, March 17, 2018, Self-driving car scenario).

4.2.

MAJOR CATEGORIES OF RISKS AND ISSUES IN RESPONSIBLE AI DEVELOPMENT

Citizens identified 12 major risk categories and issues in responsible AI development during discussions of the different scenarios. These categories are not mutually exclusive, but offer a snapshot of various themes raised by citizens in responsible AI development and warrant special attention for the purposes of creating public policies. The following mind map presents the scope and diversity of the issues discussed, which have been classified into categories and subcategories. Sometimes, dilemmas or marked oppositions came out of the discussions. The following section provides a definition for each category, illustrated with examples taken verbatim.

Table 3: Mind map of issues



GOVERNANCE

COLLECTIVISM VERSUS INDIVIDUALISM

This category refers to a dilemma which pits protecting individual interests, choices or responsibilities against protecting collective interests, choices or responsibilities. The answer to this dilemma is an important issue that strongly depends on a normative position for which no consensus was reached.

"Ensuring that AI technology is a learning tool that serves the social and democratic ambitions of school as a public good."

(Education table, SAT, Montréal, March 13, 2018, Nao scenario).

"Digital twins: this is a very libertarian way of proceeding, which once again creates tension between individual and collective well-being."

"We are at a point in democratic life where the focus on the individual is so great that it will lead to a dictatorship."

"How can we ensure that self-driving cars maximize well-being? The sharing of public spaces? How can we reconcile the safety of the majority versus that of the individual?"

"Can public interests align with private personal interests and remain ethical?"

GOVERNANCE: PUBLIC VERSUS PRIVATE

Issues related how managing AI development would be divided between **public and private institutions,** and the inherent risks were also raised. These challenges were often presented as questions: How would this be shared equitably? Which of the two methods of governance is the most appropriate?

"Who is steering all of this? What powers will the organization or company hold over this tool? Will we be dependent on the company? If it becomes a national priority, what choices will be made for educational programs when it is implemented? Is it public? Private? The entire education ecosystem will be redefined."

More specifically, the **risks of conflicts of interests, commodification of personal data or the emergence of a monopoly were raised.** Participants particularly highlighted the risk of a conflict between private interests (essentially financial) and other interests, which could limit the independence of certain stakeholders or public institutions. The risk of commodifying personal data refers to issues related to the market value of data, the limitations of collecting data and the profits associated with it, particularly with respect to the protection of privacy. The emergence of a private monopoly in the governance of AI development was also a subject of concern.

"Avoid commercial use or interests that aren't educational when it comes to data collected and analyzed by AlterEgo"

(Education table, Bibliothèque Père-Ambroise, Montréal, March 3, 2018, AlterEgo scenario).

"How to avoid excessive commodification of data and people without their knowledge?"

(Smart city and connected objects, SAT, Montréal, March 13, 2018, Smart toy scenario).

"Excessive concentration of power (GAFAM), which prevents:

- Equitable sharing of AI benefits

- The arrival of new stakeholders (new business models, e.g. co-op)"

(Workplace table, SAT, Montréal, March 13, 2018, Socially responsible restructuring scenario).

DEMOCRATIC GOVERNANCE

Given that the discussion on governance often pits public institutions against private companies, issues on another alternative were raised: that of a participatory governance which involves citizens directly. These issues include the shared and collective management of open-access digital goods **(digital commons)** and the role of **citizen involvement** in current and upcoming governance (whether present or absent).

"Issue 3: Participatory democracy with a balance of power (states, social partners, businesses, unions, etc.)"

(Workplace table, Musée de la civilisation, Québec City, April 6, 2018, Socially responsible restructuring scenario).

Citizens recognized that the urgency of the situation and a certain technological determinism were factors that could harm participatory governance. The lack of time that would eliminate any possibility of a democratic process needs to be recognized.

"Urgency instead of taking the time to hold an informed and participatory democratic debate"

(Workplace table, SAT, Montréal, March 13, 2018, Socially responsible restructuring scenario).

SOCIAL JUSTICE

Citizens brought up different risks and issues regarding algorithms biases, access to AI and the consequent discrimination or exclusion of certain groups of individuals. They considered the impact of these risks on diversity and equity to be important issues.

"Implementing nondiscriminatory algorithms that foster diversity, inclusion and social justice"

(Workplace table , Musée de la civilisation, Québec City, April 6 2018, AI as a compulsory step to employment scenario).

Accessibility issues included how to guarantee access to AI and its uses. They are associated with restricting access of certain groups or social classes. Discussions were also held on the impartiality of algorithmic systems and their potential for discriminatory bias, namely data on which the algorithms are trained, as well as data collection or even the code itself.

"The values of justice (independence, impartiality, equity) prevail over technique when deploying these tools."

(Legal system and predictive policing table, SAT, Montréal, March 13, 2018, Parole scenario).

Citizens pointed out the discrimination that could arise if the first two categories of issues (accessibility and exclusion) are not adequately addressed: the discriminating effects of AI systems, whether by reinforcing existing discrimination

(e.g. gender or social status), or creating new discrimination (e.g. people who are not "connected"). **Discrimination** issues are closely tied to the risk of **exclusion** for some people, whether they voluntarily refuse to take part in the "digital society", or whether they are involuntarily excluded.

"What happens to people who don't have a digital profile? Are they at a disadvantage? Should we rely solely on AI for recruitment? Can AI truly grasp the hiring criteria? Do we have a choice if everyone else is doing it? And how do you evaluate a digital reputation?"

These risks led the participants to identify a protection issue for:

1. The **diversity** of intelligence, skills, individuals and society as a whole.

"Does AI simply reproduce the same intelligence that is taught in school? Wouldn't it be more beneficial to cultivate different types of intelligence?"

2. **Equity** so that AI operations led to decisions and recommendations.

"Sharing the benefits of AI (productivity gains). Equity between social groups, territories (cities and regions), taking vulnerabilities into account."

(Workplace table, Musée de la civilisation, Québec City, April 6, 2018, Socially responsible restructuring scenario).

FREEDOMS

This category refers to issues of maintaining individual freedoms, especially when it comes to freedom of choice—whether being able to make your own decision when faced with an AI-guided decision, or being able to choose not to use those tools without being socially excluded (which means that these issues are often closely tied to the previous category).

SELF-DETERMINATION

Citizens discussed the risk of algorithmic systems being overwhelmingly deterministic, particularly with regard to an individual's capacity for self-determination (as opposed to a risk of blind faith in technology).

"What concerns me the most is that the grandmother is excluded from the thought process. A robot nurse, fine, but what does the grandmother want? We have to ask people what they want."

FREEDOM OF CHOICE

Being able to make individual choices as well as the right to refuse to use technology or take part in a data collection system were also discussed.

"How can we ensure that an individual maintains their freedom to choose and doesn't become a slave to technology?"

"If we need everyone's data to create collective well-being, do we have to force everyone to share their data? And if some people refuse to do so, what impact will that have on the system? This is a societal choice that must be made."

STANDARDIZATION OF SOCIETY

The standardization of society addresses risk issues that arise when AI categorizes individuals for predictive purposes in healthcare, education, justice or mobility. This could lead to individuals being stigmatized and behaviours normalized instead of encouraging diversity.

"Risk of a standard profile (normalizing behaviours)"

(Smart city and connected objects table, INM, February 18, 2018, Connected refrigerator scenario).

SOCIO-DIGITAL CHANGES

This category refers to discussions and issues on social and societal changes that could result from AI development. These changes may (or may not) lead to a true "digital transition".

ACCEPTABILITY

Citizens repeatedly brought up the issues of acceptability and social buy-in when implementing AI. These discussions revolved around issues such as maintaining the public's trust in technology (AI) and in the different sectors that might use it. They also brought up issues of technological expectations and "technophobia". At times, there seemed to be a certain sense of fatalism, particularly toward technological determinism and a somewhat forced acceptance of AI development. The legitimacy of using AI in certain fields was sometimes questioned.

"Maintaining and promoting the population's trust in the justice system"

(Legal system and predictive policing table, Musée de la Civilisation, Québec City, April 6 2018, Parole scenario).

HUMAN SKILLS

Participants repeatedly discussed the impact of AI development on human skills. For example, they deliberated the transformation of human skills from the perspective of consequences (mainly negative) that AI development could have on knowledge and abilities.

"Fear of exceeding humans, human ability to be at 360° (whereas AI has excellent, very specific skills)."

"How can we ensure that dialogue with the patient is maintained (human contact) and that the doctor doesn't lose their expertise and independence?" (Healthcare table, Bibliothèque de Sainte-Julie, March 25, 2018, Intelligent hospital scenario).

A risk of dependence on technology (and more specifically, in this case, the use of AI) was brought up.

"We become dependent (on technology)"

"AI causes us become too specialized and takes us further away from general knowledge and independent learning."

The digital literacy issues refer to the need to educate the population on AI practices and issues, so people gain both the technical and critical skills required to function both as a worker and citizen in a digital society in transition.

"To guarantee that a device like AlterEgo is used intelligently, it is important that youth, parents and teachers be made aware of how the collected data is used. This raises a knowledge issue that entails an AI literacy approach."

AI SKILLS

Regarding AI skills, issues about the true advantages led to discussions questioning the potential benefits or uses of AI.

"How can we ensure that our AI tools respect the fundamental principles of our justice system?"

(Legal system and predictive policing table, Musée de la civilisation, Québec City, April 6, 2018, Parole scenario).

"Does AI fulfill its role of improving and providing access to the health and living standards of individuals/communities (rationalization, dehumanization of patient care, unexpected effects and actual efficiency of algorithms, etc.)?"

(Healthcare table, SAT, Montréal, March 13, 2018, Digital twins scenario).

Ensuring the efficiency and validity of AI, meaning the relevance of its use and skills, was also identified as an issue.

"We have to guarantee healthcare recommendations based on:

1. algorithms that are managed, validated, updated (based on scientific knowledge) and uncompromised (security/hacking);
2. complete, honest and unbiased data."

(Healthcare table, Benny Library, Montréal, March 18, 2018, Digital twins scenario).

"If AI draws wrong conclusions, how can we ensure that we are evaluating its performance? Inevitably, AI will evolve, and we will have to plan for mechanisms to validate the results and plan for continuous evaluation."

"Yes, after every decision there must be an evaluation of that decision. If we do not evaluate the performance and consequences of decisions made by the algorithm and we continue to use the algorithm, the AI will wind up basing itself on mistakes."

The risk of replacing humans was also brought up on many occasions, and was linked to the role of AI and the duties it could perform instead of a human, the advantages and inconveniences of its use as well as the way to share skills between humans and AI.

"AI will fill certain gaps in the education system, but is it the solution? Teachers' workloads will be considerably lightened, which gives them a break, but also raises the question of replacement."

More nuanced discussions highlighted issues of a balance between the benefits and the risks of AI and its skills, or the need to take these benefits and risks into account for responsible development.

"How can we implement AI into everyday objects while harmoniously developing society (cultural aspect, well-being, child development, candour) and living beings?"

(Smart city and connected objects table, SAT, Montréal, March 13, 2018, Smart toy scenario).

HUMAN-AI SYNERGY

This category refers to discussions about the advantages of human-AI synergy or the inconveniences of such a "collaboration". The main point of discussion was the synergy between the objectivity and systemization of AI on the one hand, and the subjectivity and empathic contextualization of humans on the other.

"Ensuring AI-teacher complementarity in terms of expertise and relationships with students."

(Education table, SAT, Montréal, March 13, 2018, AlterEgo scenario).

"How can we ensure that healthcare decisions aren't solely based on objective data but also consider the context and the user's choice?"

(Healthcare table, Bibliothèque du Boisé, Montréal, March 17, 2018, Healthcare insurance scenario).

"Objective justice from AI predictions versus subjective intelligence (based on experience)"

RESPECTING HUMANS

Respecting nature and the human condition were other issues raised by the citizens. These discussions led to questions of what defines a human being, what will be left of human beings, or how to put people first in the context of AI development and the importance it could take on.

"What is a human being? What are we keeping of human beings? What do we want to keep of human beings?"

The risk of the dehumanization of activities and services with AI development or the emergence of a new form of isolation—specifically caused by decreasing socialization, or delegating social relationships to robots—were also brought up repeatedly.

"The human aspect of care is lacking. The relationship between healthcare professionals and patients"

"How can we ensure human dignity and the place of human beings in the justice system?"

(Legal system and predictive policing table, Musée de la civilisation, Québec City, April 6, 2018, Parole scenario).

"This will also result in cases being standardized, and people themselves won't be sufficiently taken into consideration."

"Relationships with AI to the detriment of humans leads to growing solitude."

TRANSFORMING ACTIVITIES

This category refers to discussions surrounding societal changes that would come with AI development and the eventual digital transition in the various sectors concerned, at different levels (for example, AI transforms knowledge, the city, the conception of work, etc.)

"We're rationalizing health."

"Redefining/transforming the nature of teacher-student relationship in a learning environment and changing our relationship to knowledge"

(Education table, SAT, Montréal, March 13, 2018, Nao scenario).

"Will increasing mental capacity through transhumanism make education obsolete?"

"There's a risk of crystallizing law. The more decisions AI makes in a certain direction, the more likely it will be to rule the same way going forward."

"In 30 years, people will sleep, work, etc. in their car, which will cease to be a device used solely for transportation. Mobility will take on a whole new meaning."

PRIVACY AND CONFIDENTIALITY

ANONYMITY, CONFIDENTIALITY AND THE BENEVOLENCE DILEMMA

This category refers to respecting **anonymity** and **confidentiality issues**. Discussions were held around the real possibility of respecting anonymity with responsible AI development, how to ensure that "sensitive" data remains confidential, or how to restrict its access to certain people and uses that would be more justified than others. At times, AI was considered the problem; at other times, the solution to this type of issue. A **dilemma** surfaced on many occasions, especially in the field of healthcare. The dilemma highlighted the opposition between benevolence (which supposes collecting as much data as possible, and not just objectifiable data to ensure a more human and context-based approach by AI), and the respect of privacy and confidentiality (which would be challenged by this very data collection).

"Confidentiality no longer exists, it's a myth. We tried making data anonymous, it doesn't work. Now we can impose that only algorithms can see the data, not the human stakeholders that handle the data."

RIGHT TO BE FORGOTTEN

Discussions were also held on creating a right to be forgotten (being able to erase personal data), and the issues and impact of implementing it.

"Right to be forgotten (storage limitation), right to modification, right to suppression"

(Legal system and predictive policing table, Bibliothèque Père-Ambroise, Montréal, March 3, 2018, Preventive arrest scenario).

INTRUSION

Discussions about the risks of intrusion into people's private lives, breach of privacy and ways to guarantee protection were held on many occasions.

"How can we ensure that various components of private life (omission, property, consent, portability) are respected in the context of connected object use?"

(Smart city and connected objects table, SAT, Montréal, March 13, 2018, Smart toy scenario).

OWNERSHIP OF PERSONAL DATA

This category refers to issues related to ownership of personal data, its definition, the consequences of this ownership on privacy (to what extent does an individual own and remain the owner of their own data?) and the protection of people's "digital reputation".

"Data concerning private life should be the property of the people concerned and shared according to rules voted on democratically."

(Healthcare table, INM, Montréal, February 18, 2018, Digital twins scenario).

SURVEILLANCE

Issues of surveillance are linked to data accessibility and profiling, which raises concerns about (constant) mass surveillance of individuals that risks violating both privacy and individual liberties.

"How can we live healthy lives when we are constantly being watched?"

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

"Could a higher power, government or company, take control of my vehicle?"

FREE AND INFORMED CONSENT

Discussions were also held on the capacity to consent to the use of AI and personal data.

FREE CONSENT

At issue was the true independence of individuals and their right to share (personal) data or not, to have a real impact on how it is managed or choose how it will be reused.

"Are we truly free to not share our data?"

"If we're sharing publicly, are we truly consenting to that information being reused?"

INFORMED CONSENT

The issue here ties into information mechanisms needed for individuals to consent in an informed manner; it concerns access to information and understanding this information. This issue is closely linked to citizens' digital literacy as well as transparency.

"The question of informed consent (for both students and parents) lies at the heart of issues of data collection and interpretation as well as student autonomy."

ENVIRONMENT/ECOLOGY

These issues concern the impact of responsible AI use and development on the environment, as well as its energy costs.

"We don't often talk about the environmental aspect: storing data, stockpiling outrageous amounts of data and the inherent energy costs."

INFLUENCES

These issues refer to concerns about AI's influence (whether undue or not), or potential for manipulation. To maintain a certain freedom in the choices guided by AI and avoid placing blind trust in these devices, citizens recognized the need to cultivate critical thinking among individuals who use AI.

LOBBYING

Citizens worried about AI creating a new type of lobbyism, which could yield too much power and influence over the healthcare system, connected objects or self-driving vehicles.

"Should it be up to politics to determine which algorithm will be used? What about a lobby for algorithm designers?"

MANIPULATION

Participants worried about the risk of users being manipulated as actions and decisions become increasingly influenced by AI mechanisms, whether unknowingly or through more explicit incentives.

"To what extent can a machine influence our decisions? Do we know what impact a connected refrigerator's suggestions will have on our daily lives?"

"Insidious influence on our behaviours without us asking for it or accepting it"

"Influence risks: How can we make the risk of influences (consumption, judgment) linked to connected object use visible? How can we ensure everyone's interests (consumers, citizens, companies) are respected? Who determines the guidelines for developing these (eco) systems, and how do they go about it?"

(Smart city and connected objects table, SAT, Montréal, March 13, 2018, Smart toy scenario).

PATERNALISM

Exposure to various forms of paternalism and control (from companies, the State) was mentioned on more than one occasion. It could be increased through incentive systems, but also through the depersonalization of relationships (namely patient-heathcare provider relationships).

VULNERABILITY

Citizens recognized that not everyone is as vulnerable to the influence risks presented. Special protection of those who are most vulnerable was highlighted as an important issue.

SHARING RESPONSIBILITY

This category refers to issues of shared responsibility in responsible AI development and the consequences of decision-making.

DISEMPOWERMENT

Disempowerment here refers to concerns about the risk of disempowerment in AI development, which could translate into delegating this responsibility to algorithms (considering their growing autonomy or the perception of a growing autonomy).

"Risk of disempowering the teacher who would defer to 'diagnosis syndrome', combined with the risk of reinforcing a certain student profile."

(Education table, Bibliothèque Père-Ambroise, Montréal, March 3, 2018, AlterEgo scenario).

"It creates a lack of accountability: say I'm hyperactive, the machine confirms it, so I put in less effort. But you have to be part of the solution, buddy. The way of working will change. A teacher's duties are going to change, that's for sure."

"Knowledge is tied to responsibility. There's a risk of disempowerment if there is a loss of knowledge. A loss of critical thinking from judges and other people."

"How can we ensure that AI remains a service and that the various stakeholders (individuals, programmers, society, etc.) aren't disempowered, remain vigilant, and that individuals are always in control?"

(Smart city and connected objects table, Musée de la civilisation, Québec City, April 6, 2018, Connected refrigerator and Carbon footprint scenarios).

ACCOUNTABILITY

This issue is about identifying who is responsible or accountable in various situations concerning AI development (the user, the developer, the algorithm, etc.?).

"Who holds the learning data, who uses it, for how long? Who is protecting it?"

(Education table, Bibliothèque de Sainte-Julie, March 25, 2018, Nao scenario).

"Who is steering all of this? What power does the organization or company hold over this tool? Will we depend on this company? If it becomes a national priority, what choices will be made for educational programs when it is implemented? Is it public? Private? The entire educational ecosystem will be redefined."

"Who manages the algorithm, who controls it, who supervises the person programming it?"

SHARED RESPONSIBILITY

Discussions were also held on sharing responsibility in AI development, the complexity of this sharing and the need to take all responsibilities and stakeholders into consideration.

"The issue of the individual and shared responsibilities, which may be conflicting, of various stakeholders (governments, healthcare professionals, patients, private companies, researchers and managers, etc.)."

(Healthcare table, SAT, Montréal, March 13, 2018, Vigilo scenario).

"Issue 2: Define everyone's roles and responsibilities (institutions, students, teachers) to provide a framework for implementing AI"

(Education table, SAT, Montréal, March 13, 2018, Nao scenario).

"I don't know of any teachers that shirk their responsibilities towards their students. But we need to involve as many people as possible, adopt a multidisciplinary approach. Not make the teacher the sole person responsible for AI or AI diagnoses. Ensure that using AI for educational purposes is a shared responsibility."

DECISION SOVEREIGNTY

Issues of decision sovereignty echo the normative expectations detailed in the recommendations ("Main anticipated directions") which state that AI must remain a tool, an assistant or an additional information resource. These recommendations were made following discussions on issues of decision sovereignty; that is, whether humans or AI should have the last word.

"Algorithms should always give advice, not make decisions. The absence of human moderation is problematic, as algorithms don't take all aspects of an individual into consideration."

(Healthcare, Bibliothèque Père Ambroise, Montréal, March 3, 2018, Digital twins scenario).

"The problem with interpreting Alterego's diagnosis is that we can't forget that human intervention is necessary. We can't rely solely on a machine."

"We delegate a lot of micro-decisions to AI and interconnected systems, at the expense of humans."

STRESS—ALARMISM—ANXIETY

Participants worry that AI development will induce stress, alarm or anxiety due to information and notification overload or a lack of human contact, among others.

"How will students develop academic independence and learn to manage their stress and emotions when they no longer have access to AlterEgo during their post-secondary education?"

(Education table, Benny Library, Montréal, March 18, 2018, AlterEgo scenario).

"We must guarantee an individual's well-being when informing and treating them: not be alarmist."

(Healthcare table, Benny Library, Montréal, March 18, 2018, Digital twins scenario).

SYSTEM SECURITY AND INTEGRITY

AI and data system **reliability** issues were discussed at several levels: validity, infallibility and robustness, the integrity of systems and the people managing them. System vulnerability (bugs, errors, etc.) and impact of breaches on different reliability parameters were also raised. The risk of system **outages** and managing these risks were also among the issues brought up. These issues are closely linked to AI skills and biases. Citizens worried about risks of **piracy or sabotage** of algorithms and collected data, whether or not it was intentional, and the risks associated with potential **misuse** of data and algorithms (without necessarily amounting to piracy) and the problems it could cause.

"I don't want to be judged later on for things I did in the past."

"What if a hacker took control of the educational development of certain students? Or if parents could have an even greater impact on their children's grades? The hacker or the parents could choose the content, and therefore how AlterEgo interprets the data. For example, parents who do not want their child pursue a career in the arts could use AlterEgo to these ends."

Intent and malice in problematic or unsecured use of AI were identified as important parameters. Citizens pointed out that it was difficult to differentiate a malicious act from a problematic act that had good intentions, and the consequences of this distinction.

"Even with good intentions, we can cause problems (inaccurate model)."

"How can we distinguish temporary behaviour with no harmful intent from a genuine decision to carry out a crime?"

On many occasions, discussions revolved around a **zero risk** possibility, and whether it was truly desirable.

"Should the zero accident objective be reached at all costs? Is this objective really worth it?"

A number of dilemmas were identified during discussions on protection of security:

> Transparency (guaranteeing transparency could increase risks of piracy)

> Efficiency (ensuring the greatest possible security involves a compromise with system efficiency, as it must be secure without becoming inoperative)

> Respect of privacy and individual freedoms (in the specific case of preventive arrests, which impose surveillance in the name of public safety)

TRANSPARENCY

The issue of transparency was formulated as the ability to understand an algorithmic decision and to react to it, whether as an ordinary citizen or a professional using AI for their job.

EXPLAINABILITY AND UNDERSTANDING

These issues pertain to the explainability of a decision and the "black box", the importance of showing the process that leads AI to a result or the intelligibility of information and the importance of it being explainable.

"Transparency of the variables used, data, parameters. Explaining a decision in plain language."

(Workplace table, Mordecai-Richler Public Library, Montréal, March 10, 2018, AI as compulsory step to employment scenario).

"The complexity of the world of algorithms does not allow us to understand how AI proceeded (...) We don't require that much transparency from judges, so why should we request as much from the algorithm?"

RIGHT TO INFORMATION VERSUS RIGHT NOT TO KNOW.

This dilemma surfaced particularly in the healthcare sector and sets the right not to know (the entire range of diagnostic predictions provided by AI, for example) against to the right to know (to respect a patient's autonomy and consent). The right not to know could be justified in the name of benevolence (if certain recommendations are alarmist and uncertain).

5. POTENTIAL SOLUTIONS AND FRAMEWORK FOR RESPONSIBLE AI DEVELOPMENT

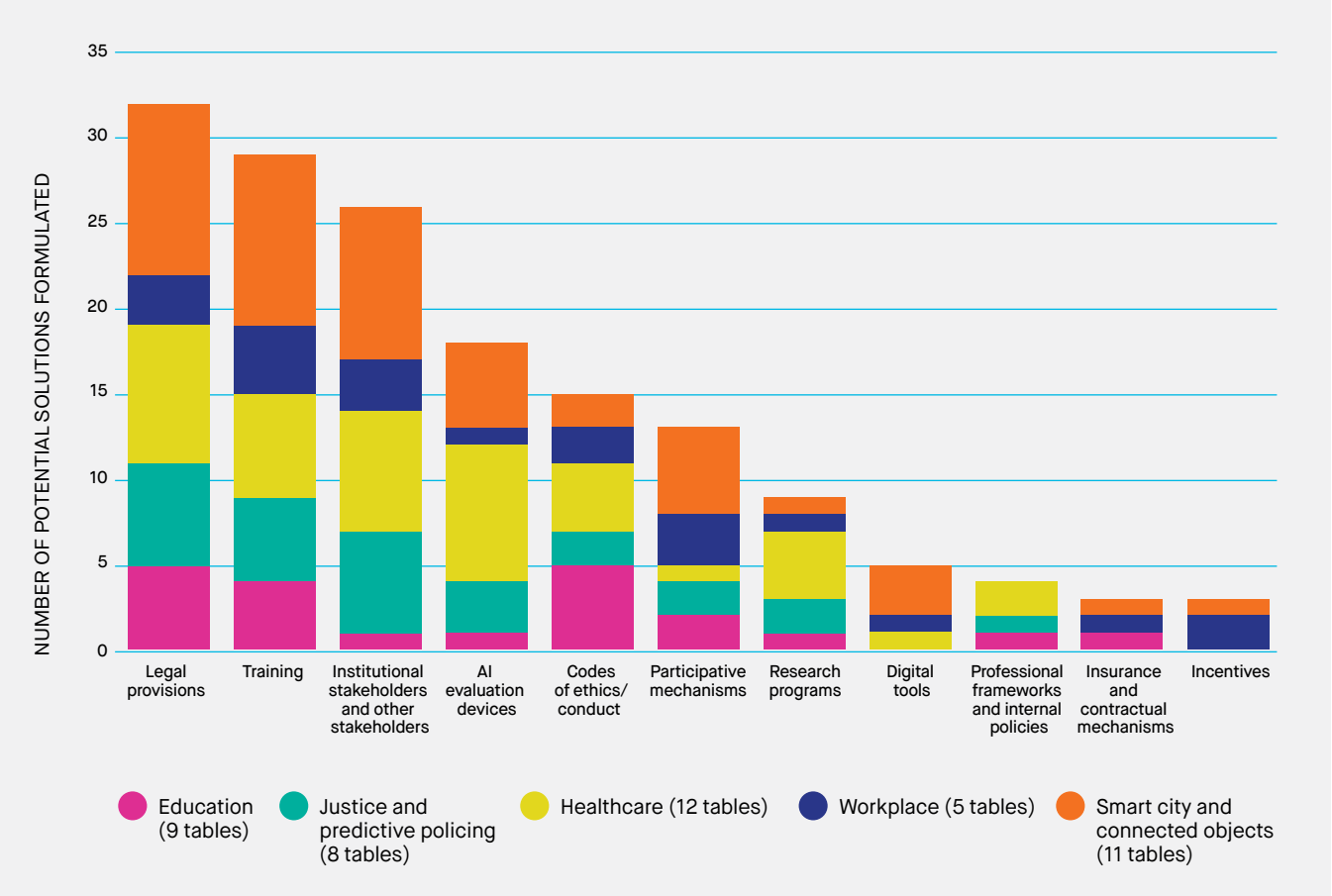
5.1.

INTRODUCTION

Citizens who took part in the co-construction days were invited to propose solutions to the previously identified issues. A total of 190 potential solutions were formulated and adopted by consensus during these activities (although other suggestions may have been discussed during the tables). By potential solutions, we mean concrete mechanisms that citizens put forward to respond to the previously identified issues.

Only possible solutions written on posters were counted. However, other recommendations were discussed or suggested (during the drafting of headlines and leads or in discussions). For the sake of coherency and feasibility, they were not included in the total number of recommendations, but were considered and analyzed when writing this section.

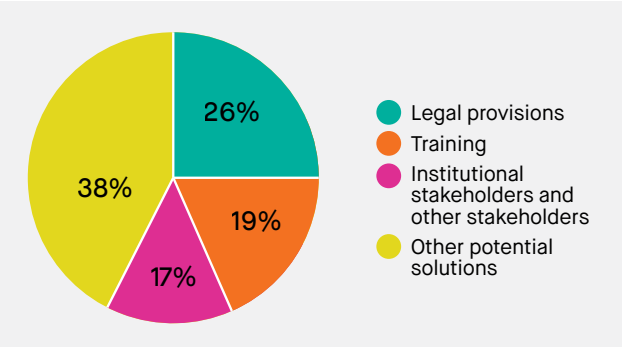
Table 1: Potential solutions proposed to respond to the issues identified



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

1. Legal provisions
2. Training for everyone
3. Identifying independent key actors for AI management

Table 4: Three key potential solutions at all tables



Regardless of the sector, all tables agreed on recommending that a legal framework adapted to the reality of AI development and personal data management (especially big data) be implemented. These restrictive provisions all refer to rights or laws. They could be laws and regulations, defending new fundamental rights, or even public policies (ranging from implementing social programs and a charter to creating digital citizenship).

Implementing training that was accessible to all was also strongly recommended, both for professionals in the affected sectors (to ensure adequate use of AI systems in their work) and the general population (to ensure everyone can participate in the debate and acquire basic digital literacy).

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

By recommending these three main mechanisms as potential solutions, a distinct trend emerges in the position held by Quebec citizens who took part in AI governance activities: it should primarily be handled by the State. Indeed, implementing incentives for businesses, or insurance and contractual mechanisms that correspond to a more liberal management were the least recommended potential solutions. These recommendations are nonetheless coherent and instructive. Citizens at different tables agreed on developing incentives—to encourage responsible development—and implementing diversity quotas (which reward companies that guarantee not to exclude or discriminate against certain minorities through AI biases) or funding for companies that help employees transition when jobs are being replaced by AI. Creating contracts between the various AI development stakeholders and its users, or insurance mechanisms to guarantee the protection of individuals in the advent of AI development was also suggested.

In all sectors, citizens suggested creating technical and ethical evaluation mechanisms for AI. Establishing a certification (or label) system as an ethical guarantee was suggested on many occasions, in particular. Different tables also recommended implementing a code of ethics (whether updating the existing code or creating new ones) and participatory mechanisms (e.g. co-constructions or public consultations) to guarantee that AI development and management remained democratic. Establishing professional frameworks (and different internal procedures for companies and institutions) that were not codes of ethics were also discussed.

The importance of implementing research programs in various disciplines (e.g. philosophy, social sciences, bioethics) to cultivate new knowledge and create digital tools (e.g. digital and interactive healthcare consent forms, personal digital file in the workplace sector) was also raised.

The following sections present the potential solutions formulated by citizens per fields of AI application. These potential solutions, defined

through concrete mechanisms, were not all discussed and developed to the same degree. Although it is evident that it is hard to imagine implementing all these recommendations given their diverse and somewhat contradictory character, an comprehensive presentation does offer, however, an especially robust global vision of the variety of solutions considered by citizens in AI management.

5.2.

EDUCATION

Table 5: Potential solutions or general guidelines for the education sector

	Number of potential solutions formulated
Legal provisions	8
Training	7
Codes of ethics/conduct	5
Participative mechanisms	2
Institutional stakeholders and other stakeholders	1
AI evaluation devices	1
Research programs	1
Professional frameworks and internal policies	1
Insurance and contractual mechanisms	1
Total	27

LEGAL PROVISIONS

Participants raised the need for creating and tightening certain laws in AI development in education. For example, a right to be forgotten was recommended regarding data use, as was an "expiration date", and no default sharing with other services unless there is a serious reason to do so. The right to be forgotten was often identified as the need to create a "data destruction policy" to allow students to reshape their identities and grow as individuals. The need to reinforce protection of privacy (particularly when it comes to data from youth) and transparency concerning data collection (namely by encouraging formats that are easily understood by users) was also brought up. For participants, a legal framework where "under no circumstance should the use of artificial intelligence limit a user's future possibilities, whether social, economic, etc." (INM table, Montréal, February 18, 2018, AlterEgo scenario), should be implemented.

Other initiatives were also formulated, such as creating a rule so that parents and students can choose to use AI devices or not, defining industry involvement in the education system to ensure ethical use of AI, and finally planning for strategies (through public policies) that would avoid "education hacking" by keeping data encrypted.

Furthermore, some citizens suggested creating a law or regulation that aims to "develop a common language (inspired by healthcare with food nutrition labels on processed foods) to bridge the gap between technology and its users" (Laval Library table, April 24, 2018, Nao scenario).

TRAINING

With regard to education, participants recognized the need to be proactive in implementing training for the entire community affected by AI development in this sector. Training would cover digital literacy, media literacy, as well as ethics and issues related to integrating AI in an educational environment. This could be, for example, digital literacy training for both parents and students, or it could be directly integrated into initial citizen training.

Citizens also recommended training education professionals specifically, for instance by including the development of work skills "in tandem" with AI devices in the curriculum for the basic and university training of teachers (e.g. a certification for the B.Sc. or an accreditation system). This training would be technology-based (how to use AI), but also geared towards teaching techniques with AI (how to organize teaching plans and emphasize that knowledgeable professionals direct AI, not the other way around).

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

(SAT table, Montréal, March 13, 2018, AlterEgo scenario).

The importance of establishing adequate training was also raised. The training's purpose would be to provide stakeholders with appropriate information to assume responsibility for AI, and to discourage teachers from putting blind faith in educational AI devices. This training would help accelerate stakeholders' understanding in learning environments and mobilize them to develop AI so that learners became autonomous and equipped to deal with these realities. This training will help develop human skill sets and provide impetus to guide and even redefine future AI development.

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

(SAT table, Montréal, March 13, 2018, Nao scenario).

CODES OF ETHICS/CONDUCT

Citizens also recommended implementing codes of professional conduct or ethics for teachers, which would focus on different ethical principles (e.g. justice) for AI use in an educational environment. These codes would provide a professional framework to prevent teachers from becoming disengaged as well as the risks of harmful use, profiling or discrimination.

> "Ensure that AI use is a shared educational responsibility (support staff, family, teachers, robot)"

(SAT table, Montréal, March 13, 2018, AlterEgo scenario).

> "Teaching while preserving the relational and emotional quality of human interaction."

(Bibliothèque de Sainte-Julie table, March 25, 2018, Nao scenario).

PARTICIPATORY MECHANISMS

Citizens suggested establishing open-source AI communities in public libraries to crack the AI "black box". The idea of leading general assemblies through consultations on socially responsible development of AI in education was also suggested.

> "Consultation in the field of education to assess the current situation and define the roles and responsibilities of each player"

(SAT table, Montréal, March 13, 2018, Nao scenario).

INSTITUTIONAL STAKEHOLDERS AND OTHER STAKEHOLDERS

Citizens suggested creating a permanent Quebec multi-stakeholder committee that would be made up not only of department officials, but also representatives for parents, students, teachers, librarians and researchers. This would be a space for public debate and would serve as a counterbalance to private companies. This committee's mandate would be to advise the government (binding recommendations); prepare codes of ethics and training; introduce and oversee open source licences and consult with citizens. Citizens also recommended setting up ethics committees that would conduct consultation processes at every level of a technology's evolution, while ensuring its social acceptability. The idea of creating a joint, inclusive and diversified committee made up of educational stakeholders was also suggested. Citizens felt that a department should be responsible for creating this committee. Lastly, certain participants recommended creating a "Department of technological access and integration for training and certifications" (Bibliothèque de Laval, March 24, 2018, Nao scenario)

AI EVALUATION MECHANISMS

Participants felt that creating certifications was mandatory, particularly to ensure that certain standards were upheld, such as respect, conscious choice and freedom. Also, some certifications could guarantee that algorithms would not be used to replace teachers. Participants recommended tests and classroom observations to ensure this type of tool does not impede students.

RESEARCH PROGRAMS

Citizens recommended the joint or parallel development of technology and human creativity through research programs led by interdisciplinary stakeholders. These programs could focus, for example, on technology and mental health, ensure freedom of choice in using AI and safeguard human autonomy in decision-making. They also recognized the need for AI in for educational research, to intervene as early as possible in a child's learning.

PROFESSIONAL FRAMEWORKS AND INTERNAL POLICIES

Citizens believe that schools that integrate AI should do so responsibly. To this end, they recommended two potential solutions: implement incentives that encourage "schools to adopt internal policies to provide a framework for AI integration" (SAT table, Montréal, March 13, 2018, Nao scenario) or establish protocols or guides that help identify certain benchmarks to help integrate AI responsibly in schools.

INSURANCE AND CONTRACTUAL MECHANISMS

Citizens stated there must be a clear commitment to preserving the well-being of students. This commitment could be a "moral or social contract" that would have to be signed by all stakeholders. Implementing it would help "clarify the degree of responsibility in protecting student well-being" (Musée de la civilisation table, Québec City, April 7, 2018, AlterEgo scenario), but also provide teachers with the right to opt out.

5.3.

LEGAL SYSTEM AND PREDICTIVE POLICING

Table 6: Potential solutions or general guidelines for the legal system and predictive policing sector

	Number of potential solutions formulated
Legal provisions	9
Institutional stakeholders and other stakeholders	7
AI evaluation mechanisms	5
Training	5
Codes of ethics/conduct	2
Participative mechanisms	2
Research programs	2
Professional frameworks and internal policies	1
Insurance and contractual mechanisms	1
Total	34

LEGAL PROVISIONS

With regard to the legal system and predictive policing, laws and regulations on transparency must be established: private and public companies that collect criminal data must be transparent, and the decision-making processes by algorithms must be able to be explained and interpreted. Explaining the decision must come with measures that allow access to mobilized algorithms and ensure they are explainable and intelligible. As an initial transparency mechanism, many deliberation tables suggested that the AI used in the legal sector—even all public sector AI—be developed in open code, under free licence. From a legal standpoint, it's about guaranteeing "the right to full answer and defence", in particular, being able to challenge a decision by raising procedural or formal defects (Musée de la civilisation table, Québec City, April 6, 2018, Parole scenario).

This call for transparency goes hand in hand with establishing legal provisions that allow for what is considered a fundamental right to be judged by a human being to preserve procedural justice and individual sentencing, but also that the appeal process for a computer-based decision is always overseen by a human judge. Many debates revolved around conciliating human and artificial stakeholders

in this process, underlining the need for the law to adapt to a new technological reality that included AI in legal decision-making. The consensus was as follows:

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

(Musée de la civilisation table, Québec City, April 6, 2018, Parole scenario).

In the scenario for using AI for preventive policing, citizens expressed the desire to establish a "framework that allows us to go beyond and eliminate biases, discrimination and abuse of power" (SAT Table, Montréal, March 13, 2018, Predictive arrest scenario) and tighten laws around consent to ensure it is truly enlightened. They also put forward the idea of limiting public and private stakeholders' access to private data, such as "private conversations on digital platforms" (Bibliothèque du Boisé table, March 17, 2018, Preventive arrest scenario) and enforcing a "right to be forgotten, to modify and correct data as well as a right to personal access to the data collected" (Bibliothèque Père Ambroise table, March 3, 2018, Preventive arrest scenario).

CODES OF ETHICS/CONDUCT

Citizens recommended establishing a declaration of principles, a code of ethics or conduct within companies, for the various professional bodies concerned or all individuals with access to algorithms. These codes would deal with consent, confidentiality, neutrality and how to protect human diversity. They would namely mitigate the speed with which AI technologies are developed, and the possibly ungovernable character of the companies that commercialize them.

"Put the declaration of principles first: Live together harmoniously," meaning that we should "continuously review and optimize algorithms so they always serve humanity and human diversity".

(Bibliothèque Père Ambroise table, March 3, 2018, Preventive arrest scenario).

TRAINING

Participants highlighted the need for awareness campaigns to develop citizens' critical thinking on AI, their right to privacy and the sharing of their data. Learning should also include digital literacy and basic skills that must be developed in primary school. The training should ensure that citizens are aware of the programs and types of data used, that they have the knowledge and necessary tools to make educated choices and better manage the information they are sharing (e.g. as an information campaign, a public event or a discussion).

Certain tables also recommended introducing mandatory training for all high school students:

"The training would include three steps:

1. The essence of AI
2. Functions and roles of AI
3. Ethical responsibility of AI"

(INM table, Montréal, February 18, 2018, Preventive arrest scenario).

Citizens also raised the need for training professionals in the field. Namely, by recommending that the judicial council define the type of training and adopt regulations to educate judges on new technological realities, so that they understand how AI works, the ethical issues related to AI and the impact of algorithmic decision-making on individuals and professionals.

PARTICIPATORY MECHANISMS

Citizens brought up the need to hold a major public consultation prior to using AI in a legal context and implementing any type of framework. The theme "For or against AI in law" would be at its core. The goal of the consultation would be to establish specific conditions for AI development in the sector prior to implementing legal AI applications. The consultation should be ongoing and evolve with new developments.

Citizens also suggested implementing consensus-based decision-making mechanisms that could be a co-construction session involving all stakeholders (professional bodies, associations, litigants, Department of Justice, industrial sector, etc.) when AI tools are acquired and deployed. They also highlighted the need to include AI users in this sector (e.g. judges, lawyers), who must be involved when selecting the product. In short, citizens felt that there was a need for consensus-based decision-making with stakeholders during the acquisition and deployment of the tool.

RESEARCH PROGRAMS

Citizens also recommended implementing university, industry and multidisciplinary research centres or programs focusing on the social, ethical, economic and political impact of AI on our society and the lives of individuals. Participants felt it was crucial to:

"Ensure that research generates solid data about the use of AI in law."

(SAT table, Montréal, March 13, 2018, Predictive justice scenario).

INSTITUTIONAL STAKEHOLDERS AND OTHER STAKEHOLDERS

As they pondered how to adapt AI tools to respect the fundamental principles of the legal system, many participants raised the need to create an independent organization to certify AI tools. It would not be to certify the tool's decision, but rather the algorithm's decision-making process. This would help ensure that the data is free of bias and that the algorithm is transparent and interpretable. Monitoring the tool's quality should continue after certification, through an audit process, for example. Many tables suggested that these independent organizations be hybrid entities (made up of public/private stakeholders, engineers, law professionals, social science researchers, ethics philosophers, etc.).

"The purpose of this entity would be to control AI. It would identify potential biases and would be achieved through co-construction"

(SAT table, Montréal, March 13, 2018, Predictive justice scenario).

Participants also brought up the need to create an independent group or body—made up of citizens and members of society—as a recourse in the event that certain principles of fundamental rights or justice were not respected. Likewise, they suggested creating a department of data ethics and digital protection, especially to preserve diversity and live in harmony with others.

Lastly, some participants suggested creating an "Artificial intelligence centre for civilian security" (AICCS) to ensure freedom, security and justice for all. "This centre, made up of citizens and professionals" aims to control "The abusive use of AI and highlights its first role and ultimate purpose, which is to be a tool that serves citizens." (INM table, Montréal, February 18, 2018, Preventive arrest scenario).

AI EVALUATION MECHANISMS

Citizens regularly put forward the need for institutional stakeholders to create standards and introduce certifications (on the creation and training processes for algorithms) that aim to protect rights and freedoms in the age of AI. They also talked about leading multidisciplinary studies a priori and impact studies a posteriori, running tests and reviewing and updating algorithms. Some also suggested creating a certification for "clear data and explicit intentions"

(Bibliothèque Père-Ambroise table, Montréal, March 3, 2018, Preventive arrest scenario). This would be an ethical certification on data dissemination and its objectives for the corporate world and government departments, in particular.

PROFESSIONAL FRAMEWORKS AND INTERNAL POLICIES

Participants expressed concerns that companies commercializing AI would become extremely adept at avoiding any form of control. They had two recommendations to this end. First, implement an ethical procedure within companies. Second, oblige private or public companies to write a mandatory annual report on significant incidents linked to AI use, out of a concern for transparency.

INSURANCE AND CONTRACTUAL MECHANISMS

Participants expressed the need for trade secrets to be lifted for legal stakeholders and citizens. This could be accomplished by introducing contracts between industry and legal stakeholders that specified the need to make the code open, examinable and verifiable for legal stakeholders and citizens.

> "AI code should be open-source and the decision should be as explainable as possible"

(SAT table, Montréal, March 13, 2018, Predictive justice scenario).

5.4.

WORKPLACE

Table 7: Potential solutions or general guidelines for the workplace sector

	Number of potential solutions formulated
Training	8
Institutional stakeholders and other stakeholders	5
Legal provisions	7
Incentives	3
Participative mechanisms	3
Codes of ethics/conduct	2
Digital tools	1
Public policies and guidelines	1
AI evaluation mechanisms	1
Research programs	1
Total	32

TRAINING

Citizens recommended implementing workplace training for everyone so that knowledge on current AI development issues could be shared. This training would help reinforce digital literacy and individual skills, as well as guarantee that citizens and future generations are aware, trained and ready for the current digital transition.

This workplace training would need to take rapid changes and uncertainties in AI development into account. This could be achieved by upgrading school curriculums, establishing awareness or support programs by the government (e.g. digital literacy programs for adults) or ongoing training for professionals. In particular, citizens came up with the idea of government agencies establishing public training for AI and digital realities to so that every segment of the population could benefit from its development.

"A major awareness program on the transition to AI as well as support programs launched by the government"

(Musée de la civilisation table, Québec City, April 6, 2018, Responsible restructuring scenario).

To avoid challenges related to AI use in recruitment, human resources professionals should also follow rigorous training on the methodological foundations of algorithms, digital data collection and the legal framework, and biases that are present or possible in AI analysis. An accelerated upgrading process and professional programs must be created with CEGEPs, universities, government departments, and professional bodies impacted by AI (e.g. law, healthcare).

INSTITUTIONAL STAKEHOLDERS AND OTHER STAKEHOLDERS

Citizens suggested creating three types of institutional stakeholders: a Crown corporation for AI in Quebec, an interdepartmental committee that advises the premier and governance committees in all companies that use AI in their recruitment process.

The mandate of the Crown Corporation for AI in Quebec, or NSAIQ (National society for artificial intelligence in Quebec) would be to support the digital transition through public policy expertise and provide assistance to private and public organizations, while opening up a democratic dialogue for AI implementation in public services:

"Its different mandates are:

> Ensuring AI expertise for drafting public policies (work, jobs, training, land use planning, education, etc.)

> Organizing democratic testing and implementation of AI in society and public services

> Supporting public and private companies throughout the transition

> Supporting and advising ministers on social programs in Quebec

> Help Quebec in international work groups"

(Musée de la civilisation table, Québec City, April 6, 2018, Socially responsible restructuring scenario)

The suggested multiparty committee would be a permanent, joint committee on economy, jobs, education and culture (inspired by the Digital Strategy). It would act as a direct advisor to the premier. This committee would allow the government to benefit from expertise independent of consultants and would not rely on private companies or third parties.

To ensure best practices for companies in AI-assisted recruitment, the suggested governance committees would be established in every company that uses AI in its recruitment processes. The mandate of these committees (one per company) would be to ensure that the code of ethics for human resources advisors is respected (see "code of ethics"). It would also ensure ongoing training for recruiters to ensure that they remain watchful for unpredictable biases that can occur at any time, and take into account the evolving nature of AI. Each company's committee would be multidisciplinary, made up of AI experts, HR experts, and people working outside the fields of AI and HR to allow for diverse opinions and experiences, and maintain a certain independence. Implementing an AI office in companies was also suggested to allow workers to see if AI use by an employer is acceptable from a legal standpoint.

LEGAL PROVISIONS

The legal provisions suggested by participants sought to address two main issues: guaranteeing human-focused AI development with an update to the Charter of Human Rights and Freedoms, and protecting (and reviewing) personal data.

"Updating the Charter of Human Rights and Freedoms to include AI and put humans first."

(Musée de la civilisation table, Québec City, April 6, 2018, AI as a compulsory step to employment scenario)

In a legal context, the idea of company accountability was defended, particularly when it came to protection of privacy: in the event of a predictive model likely coming into conflict with the existing legal framework, the company responsible for the model should communicate the information needed to evaluate its impact. Similar to the protection of privacy, the protection of personal data at work could be ensured by a regulation requiring that users be made aware that their data is being processed, as well as what data the company has, who has access to it, for what purposes, since when and for how long. All individuals should be able to access and understand this information, which could be stored in a personal digital file (see "digital tools").

Moreover, regarding the risk of exclusion inherent to holding compromising data, participants suggest allowing a form of "digital rehabilitation" for citizens who may be unfairly judged by digital footprints. A legal framework should be drafted to guide this kind of right to be forgotten, particularly to deal with the delays and the specific nature of this digital rehabilitation. This would also allow citizens to choose what information about them is made available, namely on social media.

"We must respect the existing legal framework, especially fundamental rights that already prevent discrimination when hiring. We suggest adding the right to digital rehabilitation (or the right to be forgotten) [so people aren't unjustly sidelined for digital footprints consulted by potential employers]."

(SAT table, Montréal, March 13, 2018, AI as a compulsory step to employment scenario).

Citizens also discussed creating anti-discrimination laws for algorithms or a minimum guaranteed income to help protect jobs lost in the transition.

Participants also stressed that the law needed to adapt to the many issues compounded by AI, but remain somewhat flexible in its review process to respond to the evolution of AI and its effects. Participants also recommended an "experimental approach" to avoid introducing regulations that are destined to change quickly.

INCENTIVES

Citoyens recognized the need to implement various incentives to encourage responsible AI development in the workplace, particularly with respect to the digital transition and protecting employee well-being. They brought up the need to reconsider how society directs public funds to AI and to demand socially responsible investments.

"Directing investments towards responsible AI for the common good."

(SAT table, Montréal, March 13, 2018, Socially responsible restructuring scenario).

These investments, along with employee pension funds, would come from the State and individuals joined by public advisors in corporate social responsibility, and resemble a digital transformation fund. Companies that establish a transition process for employees whose jobs are being replaced by AI could then receive subsidies (e.g. training with measures to encourage or ensure employee loyalty once training is completed).

Along the same lines, another potential solution was to create a fund to which both companies and workers contribute, which could lead to creating digital insurance (see "insurance mechanism"). In particular, this could potentially offset job insecurity by establishing a guaranteed minimum income.

Citizens also highlighted the need to review company structures to encourage including women (cross-sector consideridations), especially if the

future of work is in this field to mitigate risks of inequality. Citizens therefore suggested that funding be based on a points system that cultivates diversity (a type of diversity quotas for businesses supported by reinforcement policies, rather than sanctions).

Lastly, citizens liked the idea of developing a support program to create new business models for data processing businesses, such as co-ops. Their purpose would be to break the isolation of self-employed individuals, whose numbers will keep increasing.

Generally speaking, out of a political concern for sharing AI benefits and to ensure equitable distribution among social groups, territories and various vulnerabilities, participants recommend developing an AI development incentive policy that ties responsibility to business subsidies.

PARTICIPATIVE MECHANISMS

Participants suggested creating a multi-sector "permanent consultation space" within the government to respond to the division of powers (tied to the democracy principle) and address the challenges of how emerging sectors are structured.

Citizens also mentioned the importance of user participation in designing AI interface tools. They could be "design thinking" with different partners, and would allow them to review the work of the programmers, particularly to correct biases:

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

(Musée de la civilisation table, Québec City, April 6, 2018, AI as a compulsory step to employment scenario).

User feedback could be given to competent authorities (e.g. ethics committees, corporations) to adapt the system.

CODES OF ETHICS/CONDUCT

Two types of codes of ethics were suggested by citizens for the theme of workplace transformation: one for human resources advisors (CHRA) so recruitment efforts are carried out in unbiased fashion, the other for any profession using personal data for marketing purposes—such as advertisers—to ensure better protection of personal data.

The first, the CHRA code of ethics, would address the issue of "cultivating diversity through team building" and would be based on the results of a research program that studies recruitment biases and measures AI's impact on them (see "Research programs").

The second code of ethics would address the issue of protecting personal data. Participants call for "society to reflect on the use of personal data" in a context where they feel that the notions of "responsibility" and "common good" should be the subject of a democratic dialogue. This code of ethics would result from this societal and democratic thinking process and could be inspired by Europe's General Data Protection Regulation (GDPR).

> "Beyond individual consent (e.g. when visiting a website), we must reflect as a society on data use and issues of wealth redistribution."

DIGITAL TOOLS

A digital tool was suggested for the workplace sector: the creation of a personal digital file. This would consist of a unique portal to our digital data that obliges every business to declare the data it collects. This type of tool would have to be developed so that it operates transparently and intelligibly, particularly when using and storing personal data.

INSURANCE AND CONTRACTUAL MECHANISMS

To guide the digital transition and its impact on the workplace, citizens suggested creating digital AI insurance to allow each individual to become familiarized with AI and receive training on it. This insurance would be financed by a fund to which both workers and companies contribute (based on the same model as Quebec's parental insurance, adapted to the worker's reality). It could even facilitate access to training, and this training would be paid by companies (with an incentive measure, or even an employee loyalty program at the end of the training). Digital insurance could also help ensure a guaranteed minimum income to counter job insecurity for at-risk workers.

AI EVALUATION DEVICE

Impact studies were suggested to ensure that humans always come first in any AI system. These would be carried out by an independent organization funded by a data tax (based on the carbon tax model).

"When analyzing and creating any system, we must guarantee and maintain monitoring through an independent third party (if necessary), to put humans first. This organization would be funded by a data tax (like a carbon tax)."

(Musée de la civilisation table, Québec City, April 6, 2018, AI as a compulsory step to employment scenario)

RESEARCH PROGRAMS

Participants recommended developing multidisciplinary research programs that measure AI's impact on recruitment biases. In particular, this research program would inspire the creation of a code of ethics for HR advisors.

5.5.

HEALTHCARE

Table 8: Potential solutions or general guidelines for the healthcare sector

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

LEGAL PROVISIONS

Several recommendations on rules and regulations were made at specific levels, particularly with regard to privacy, transparency, data collection and universal healthcare.

Many citizens felt that, although we can rely on existing laws and regulations when it comes a person's rights to control their personal data, we must also think of ways to redefine them to take technological innovations in AI into consideration. Protection of privacy was an important element in these discussions, and citizens expressed the need to guarantee the confidentiality of personal data.

"Laws should be introduced to guarantee private ownership of personal data (e.g. a law giving access to data collected from the people concerned)"

(INM table, Montréal, February 18, 2018, Digital twins scenario).

Certain tables also mentioned the need to implement laws and regulations that outlined clear and transparent objectives for collecting, using and accessing biological data (and any other personal health information). This information must be clear, understandable and readily available

to users. Participants highlighted the need to outline instructions for government organizations to provide intelligible, quality and relevant information when collecting personal health and biological data.

As to data collection, they also underlined the need to oversee sources used by the algorithm to ensure there are no biases against citizens. Participants also recommended introducing laws and regulations on the goals of the healthcare system to maintain a fair healthcare system, particularly in relation to the universal healthcare principle:

"Include all AI developments in healthcare in the law on access to universal healthcare, at the same level as alternative medicine"

(Mordecai-Richler Public Library table, Montréal, March 10, 2018, Vigilo scenario).

In the context of Canada, the suggestion was made to implement a law specifying if (and how) public healthcare coverage offered by the RAMQ could apply to technological innovations related to AI in healthcare.

Finally, participants also suggested that a regulation overseen by the College of Physicians be introduced to ensure that humans always come before AI.

"Robots must not be used without the supervision of a (human) institutional authority subject to a code of ethics."

(Mordecai-Richler Public Library table, Montréal, March 10, 2018, Vigilo scenario).

INSTITUTIONAL STAKEHOLDERS AND OTHER STAKEHOLDERS

Citizens recommended implementing many institutional committees and stakeholders in healthcare. These could be advisory committees whose mission would be to define the "values" that AI should consider when processing information. Citizens came up with the idea of creating an independent organization that could rule on privacy benefits and risks while also focusing on healthcare and ethical AI issues. Participants also felt committees needed to be established to review mistakes made by AI devices to improve algorithms. It could namely be requiring that the healthcare system periodically review the validity of its algorithms, and render public how they function and are evaluated with a "declaration of any modifications" clause". (Bibliothèque Père-Ambroise table, Montréal, March 3, 2018, Digital twins scenario).

Some participants thought someone should be designated as legally responsible so a human being is held accountable in the event of error. Likewise, a forum to appeal a decision made by an algorithm must be available. Establishing an independent ombudsman whose role would be to settle disputes between patients and doctors was also suggested.

Other citizens felt that appointing a life and well-being commissioner who "rules on healthcare objectives while defending citizens and the general population, and namely the right not to know" is crucial (Musée de la civilisation table, Québec City, April 6, 2018, Digital twins scenario). Creating a body to establish a humane and independent governance framework for AI development in healthcare was also suggested. Lastly, citizens recommended implementing a healthcare data anonymization centre managed by the government whose purpose would be to protect citizens from having their personal data misappropriated by private companies.

AI EVALUATION MECHANISMS

Citizens recommended establishing ethical AI certification in healthcare; namely developing a certification (or label) for algorithms and robots from research project databases (participative study on what influences AI development) to determine the criteria and various levels of this certification. The criteria should include transparency, security and relevance of the tool. For example, these certifications would be designed to standardize access to the algorithmic decision-making process, or validate the tools of healthcare robots. These certifications should be issued by the government

or independent, multiparty organizations to protect public interest and patient well-being, and mainly target private companies developing AI healthcare.

"Upfront certification for healthcare robots and their tool kits (particularly to protect public interests)"

(Mordecai-Richler Publid Library table, Montréal, March 10, 2018, Vigilo scenario).

TRAINING

Participants recognized the need to establish education and awareness measures for all stakeholders involved in AI development for the healthcare sector including healthcare professionals and the public. Professional training, which could be in the form of ongoing training (e.g. based on creating a best practices guide) should particularly focus on the doctor-patient-AI relationship, with case studies and updated statistics. The purpose of this training would not only to make an optimized and informed use of algorithms, but provide adequate and accurate communication of information to patients to avoid misinterpretation.

As for public training, participants recommended that awareness begin from day one of the younger generation's education (in school) to cultivate critical thinking about AI technologies. Citizens proposed the idea of an intellectual self-defence class to develop critical awareness and educate users about new practices through outreach.

"In primary school, start raising awareness among younger generations and cultivating critical thinking. Ensure information shared with the public is accurate and determine what information deserves to be shared with citizens/patients."

(Musée de la civilisation table, Québec City, April 6, 2018, Digital twins scenario).

CODES OF ETHICS/CONDUCT

Citizens also recommended adopting codes of ethics, whether for any company creating AI for healthcare, or more globally for Canadian users and healthcare professionals. These codes must contain standards as to the safety, transparency and responsibility of doctors or developers. These codes should help ensure that every citizen is accompanied by a doctor for any medical decision. Some citizens mentioned that the definition of human responsibility toward AI needed to be added to existing codes of ethics. For example, it was suggested that a Hippocratic oath 2.0 be implemented. This would ensure that people receive personalized care and monitoring by including healthcare professionals in all healthcare recommendations. This could involve implementing "virtual guardrails" to prevent the algorithm from going off-track and skewing the diagnostic.

"The doctor's responsibility and code of ethics should always prevail over AI. AI is just a tool to help."

RESEARCH PROGRAMS

Citizens recommended establishing, funding and fostering various multidisciplinary research programs on AI in healthcare. Participants all agreed that AI research should be at the forefront, but so should other disciplines that study the effects of AI on society, such as social sciences, philosophy or bioethics. These studies should, for example, help identify shared responsibilities among the various stakeholders, measure the impact of AI on their autonomy or launch training and education programs for both practitioners and citizens.

"Develop research programs to evaluate the degree to which an individual's socioeconomic status has an impact on their health and eventual AI diagnosis"

(INM table, Montréal, February 18, 2018, Digital twins scenario).

PROFESSIONAL FRAMEWORKS AND INTERNAL POLICIES

In response to the risk of attacks on privacy, citizens recommended that the healthcare system be responsible for documenting and informing patients when their data is accessed by third parties ("who" and "when").

Citizens also recommended a procedure to follow for a diagnosis (in the same vein as a combined human-machine diagnosis). This procedure would encourage doctors to make a diagnosis before the algorithm, which would help safeguard the doctor's expertise and independence, and ensure the algorithm remains a complementary tool to inform the doctor and assist them in decision-making. This algorithm would not only strictly consider a patient's medical data (e.g. biological indicators), but other kinds of data (e.g. lifestyle, eating habits).

PARTICIPATORY MECHANISMS

Citizens highlighted the need to hold a debate and public consultation on data safety before introducing one or many bills. These debates should include the public, experts and other stakeholders who are already involved (e.g. ethicists).

"We have to go beyond the context of an ordinary citizen on their computer to dealing with a privacy policy."

DIGITAL TOOLS

The creation of an electronic consent form adapted to the digital reality was suggested. It should be user-friendly, digital and interactive, and a contact person should always be available to consult.

5.6.

SMART CITIES AND CONNECTED OBJECTS

Table 9: Potential solutions or general guidelines for the smart city and connected objects sector

	Number of potential solutions formulated
Legal provisions	14
Institutional stakeholders and other stakeholders	10
Training	10
AI evaluation mechanisms	5
Participative mechanisms	5
Digital tools	3
Codes of ethics/conduct	2
Incentives	1
Research programs	1
Total	51

LEGAL PROVISIONS

Participants at tables discussing the smart cities and connected objects theme suggested implementing a number of legal provisions. The goal of these potential solutions would be to protect personal data and user consent and guarantee the loyalty of technology. For example, citizens suggested a regulation authorizing disconnection at any time as a means to control connected objects. Also, in response to various risks (including invasion of privacy), participants invited people to consider including a legal provision on the loyalty of connected objects, which would guarantee that the measures taken and the recommendations made are in the interest of the consumer, not the company:

> "Law defining the notion of loyalty and other ethical considerations (discrimination)"

(SAT table, Montréal, March 13, 2018, Smart toy scenario).

Citizens recommended legally determining an age for "digital maturity" for use of technology by minors:

"We have to think about an age for digital reasoning. About digital maturity."

This measure echoes the suggestion to cultivate "digital citizenship", which would help empower citizens to deal with changes dictated by new technologies. This would help define responsibilities and educate citizens on their rights and responsibilities regarding AI accessibility, in particular.

Citizens also came up with the idea of introducing a moratorium. It could last one or two years and would help provide a legal framework for the use of artificial intelligence in public transportation:

"Prior to implementation, we must set some parameters. We need to impose a moratorium until we have responsible technology."

For equity issues, participants suggested establishing a mobility social assistance program that would help remove barriers to AI access for certain at-risk categories of people. Likewise, citizens recommended establishing a right to mobility that ensured everyone access to transportation. Reforming transportation laws, traffic regulations and road safety was therefore suggested. Citizens also felt that urban planning laws needed to be reviewed; for example, by introducing regulations that promoted mixed development and took population diversity into account.

Establishing regulations to help secure personal data and information sharing was also recommended. These regulations would help protect anonymity and data ownership, ensure the proctection of privacy or prohibit data capture outside of planned hours. These laws should also provide for greater transparency in the handling of personal data by the private sector.

"Broaden the scope of the law on consent to guarantee that individuals maintain ownership of their own data."

(Musée de la civilisation table, Québec City, April 6, 2018, Connected refrigerator scenario).

Citizens believe that these laws need to be integrated into the Constitution of Canada. To protect users' transportation parameter choices, citizens suggested introducing federal laws while maintaining regulations that could be adapted to the local level.

INSTITUTIONAL STAKEHOLDERS AND OTHER STAKEHOLDERS

Table participants discussing the theme of smart city and connected objects came up with many ideas for creating institutional stakeholders, whether independent societies or advisory committees. The democratic ideal of committees or assemblies that allowed for citizen participation was brought up many times.

For the control of connected objects, two models were suggested, including a mechanism where private stakeholders would be forced to self-regulate:

> Based on the model of the Régie du logement du Québec, a **Régie des objets connectés** (Connected Objects Board) would help set prices for connected objects (such as refrigerators) and would provide social assistance to help people buy them. It would also issue ownership certificates when purchasing a connected object, to confirm that the data generated by this object belongs to the user. This person could then choose whether they consent to the data being communicated to the marketing and insurance company, without any risk of penalty.

> An independent data management authority would allow citizens to file a class action suit in the event of harmful use. It could also manage a digital platform where users could speak freely and publicly about the advantages

and disadvantages of AI devices and thereby have an impact on the brand image of private stakeholders marketing these devices. These private stakeholders would then be forced to self-regulate through user pressure on their brand (Musée de la civilisation table, Québec City, April 6, 2018, Connected refrigerator scenario).

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

For self-driving car regulations, the creation of the SAIAQ (Société de l'Assurance de l'Intelligence artificielle du Québec) would introduce changes to road safety laws to include autonomous driving. It would also provide auto insurance 2.0 that would offer new kinds of contracts for this type of driving (Bibliothèque du Boisé table, Montréal, March 17, Self-driving car scenario).

To organize smart city networks efficiently and optimize the urban system managed by AI, participants suggested a hybrid organization: the MAIUO (Mobility, artificial intelligence and urban optimization) funded by the Quebec government (SAT table, Montréal, March 13, 2018, Self-driving car scenario). This centre's mission would be to manage and optimize the engineering for AI and pool knowledge to help draft laws and regulations following pilot projects.

Participants also considered training different groups of people, such as a Minister of Technological Development that would advise the Minister of Smart Territories, which in turn would provide a framework for urban changes related to AI and sustainable cities; or even a commission to defend the right to mobility for self-driving vehicles to guarantee protection of the right to mobility (see Legal provisions).

TRAINING

Participants recommended implementing training for citizens on new technologies and smart cities, so they could gain a better understanding of how AI operates and the new standards that come with it. This education could be in the form of outreach, ongoing training or awareness campaigns. It could, for example, focus on AI operations and use, or civic life and the digital city.

Participants recommended collective vigilance training for responsible AI use. This training would democratize AI information to educate individuals on its rules of use, cultivate informed choices and allow them to take part in the decision-making process.

"Data literacy courses offered at different levels of education to provide citizens with the tools and reflexes to make informed choices"

(SAT table, Montréal, March 13, 2018, Smart toy scenario)

AI education in the city sector must occur at every level and in different locations (e.g. library, co-op, fab lab, school or non-profit organization). It could be a hands-on course in schools to teach students how to manage different connected objects, or digital literacy education programs.

AI EVALUATION MECHANISMS

Citizens recognized the need to implement mechanisms to evaluate the costs, side effects and impacts of AI-specific policies. They considered establishing standards (e.g. ethical labels) to protect the consumer, put human beings first in decision-making and foster inclusion. For example, citizens suggested establishing an ISO-like certification that would recognize companies that offer digital services with added value for citizens. This standard would guarantee that users' control their choice of services to prevent services from becoming intrusive.

Creating a certification that ensured collaboration between humans and machines was also suggested. It would guarantee user safety, security, operability, transparency, loyalty and/or trust:

"Certification that measures and guarantees the level of loyalty and other ethical considerations of my connected object"

(SAT table, Montréal, March 13, 2018, Smart toy scenario).

PARTICIPATORY MECHANISMS

Citizens recommended implementing public assemblies such as hybrid democratic forums so that citizens could evaluate projects and user needs, and determine how public spaces are planned according to people's needs and society's values. Citizens also suggested implementing a class action system for abusive use of AI, which would be dynamic, flexible and able to adapt to technological progress.

Other suggestions involving active citizen participation were presented, such as introducing surveys and participative planning (evaluating urban planning during the transition period), systems, even an open-source code of ethics (to find solutions to community issues and improve community well-being). Citizens highlighted the need to review jurisdiction between the province, municipalities and districts.

DIGITAL TOOLS

Participants suggested integrating a type of development into the design of connected objects that would allow users to easily understand and visualize the data generated by objects (who/when/where they are sending it and why), to ensure that they could easily customize their settings. The idea would be to ensure a multidisciplinary design of connected objects that integrates the emotional and psychological aspects of an individual's relationship with food or other elements into the design process (see Connected refrigerator scenario) or recommend travel options based on personal criteria (see Self-driving car scenario).

CODES OF ETHICS/CONDUCT

Citizens also recommended introducing a code of ethics for computer engineers and AI designers, which could be implemented and monitored by an independent organization. It would rule on the need for transparency and traceability, inclusion and factor in risks to protect the public. This code would be a responsibility permit to protect the common good.

INCENTIVES

Citizens recognized the need to establish incentives to encourage companies to reveal their sources and biases, the algorithms they use and ensure the transparency of recommendations and actions of connected objects (e.g. through tax breaks or calls for tender). These incentives (whether individual or collective) could also encourage the use of other means of transportation (see Self-driving car scenario). For example, these incentives could be a mobility points system for individuals who use shared transport, especially that which runs on green energy or has low greenhouse gas emissions.

RESEARCH PROGRAMS

Participants highlighted the need to conduct studies to understand the implications of AI use and guarantee the harmonious development of society at various levels as well as reflect on preserving human heritage.

"Conduct studies to understand the implications of AI use and guarantee the harmonious development of society (psychology, culture, social issues, equality, education"

(SAT table, Montréal, March 13, 2018, Smart toy scenario).

They also suggested establishing pilot projects that promoted public transit in the city and took social equity issues into consideration, while helping eliminate design barriers.

INSURANCE MECHANISMS

Although these recommendations did not make their way to the posters, participants suggested implementing digital insurance to ensure integrity and protect ownership of personal data, whether for self-driving cars or connected objects. Moreover, creating new types of contracts for automobile use was suggested to ensure proper AI management for individual mobility.

6. CONCLUSION

A number issues and potential solutions were identified as a result of this deliberation workshop which brought together hundreds of citizens, whether enthusiasts, users or experts. The goal was to listen to what citizens had to say about responsible AI development, and discussions were organized around scenarios that showcased the many risks and various ethical issues that had been identified ahead of time, echoing the Declaration's principles. These observations should help overcome skepticism of AI development which may emerge from these results, without necessarily ignoring it. The results give us a certain idea of the social acceptability of AI and its development.

The wide range of suggestions implies that we deepen the analysis to make recommendations for public policies. All results presented raised a number of issues, which must be analyzed further in order to formulate these recommendations. Focusing on these issues appeared crucial to issue a statement on a responsible framework for AI development. They are discussed in the following priority projects of this report:

1. Addressing the challenges of AIS **governance**
2. Developing **digital literacy** for all citizens
3. Ensuring **diversity** in AIS development
4. Promoting strong sustainability AIS development

AI development therefore raises many societal issues. Although these challenges are not all necessarily specific to AI, the transformations caused by its development in various social spheres call on us to question ourselves as citizens and on the society we wish to build. At the heart of this tension between hope and fear, it is the relationship between humans and technology that needs to be highlighted. If one request seems to be unanimous, it is ensuring that humans remain front and centre in a world that is increasingly becoming artificially intelligent.

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