Responsible Artificial Intelligence: **a guide for deliberation**

ALGORA

Université de Montréal

CIFAR

INTERNATIONAL OBSERVATORY ON THE SOCIETAL IMPACTS OF AI AND DIGITAL TECHNOLOGY

Responsible Artificial Intelligence: **a guide for deliberation**

Algorithm, data, artificial intelligence (AI): all these terms have become part of our everyday life but require some clarification. How can we understand these technologies? They are certainly very much talked about and promise a better future. But what are the risks involved if we lose control over their development?

To meet the challenges associated with accountability in AI development, many public institutions, private bodies and international organizations have published charters of practice, declarations of principles and recommendations in this respect. They have shown convergence around key principles (justice, independence and well-being). However, principles are sometimes abstract and are not always defined in the same way around the world.

Much remains to be done in this area. First, we should reflect about implementing general ethical principles, ensure that they can be applied to each area of specific activities (education, science, information, health, etc.) and to put into practice the proposals resulting from this reflection. Secondly, it is essential to involve citizens more in defining guidelines for the responsible use of AI and mass data and to gather their informed opinions. Indeed, the deployment of AI affects all of us and raises ethical and political questions that should be the subject of public deliberation. Finally, it is essential to consolidate levels of digital literacy among citizens, which translates into informing and helping citizens to better understand the issues involved around accountability in the development of AI and to participate in public deliberations on the principles and standards of its deployment.

Participation in deliberative workshops on AI and digital technologies is based on a good understanding of the ethical and societal issues of AI and the rules of deliberation. The purpose of this guide is precisely to make AI and related ethical issues accessible, and to provide an introduction to deliberation on the ethics of AI. The guide includes definitions, illustrations and case studies. In this way, it creates a common language around the ethics of AI. This guide also aims to equip communities to organize their own deliberations on AI and the deployment of digital technologies in their social environment. It is designed to enable teachers, community representatives, citizens, and administrators to bring the debate to life and define common strategies.

This deliberative and participatory approach is based on confidence in people's ability to design their future and the kind of society in which they wish to live, to formulate the ethical and political principles that should organize it, and to develop relevant public policy proposals.

Finally, this document has some unavoidable limitations. It is intended to be simple for clarity and efficiency, but it is also culturally tinted. For this reason, it will be adapted to the different geographical and cultural realities in which deliberations will take place. Everyone is invited to enrich it. It is our hope that this guide will promote deliberation among citizens, stakeholders and those in charge of public affairs, and that the workshops and deliberation forums it facilitates will contribute to a more accountable and democratic development of AI.

The Algora Lab team,

Part 1
Artificial Intelligence

WHAT IS AI?

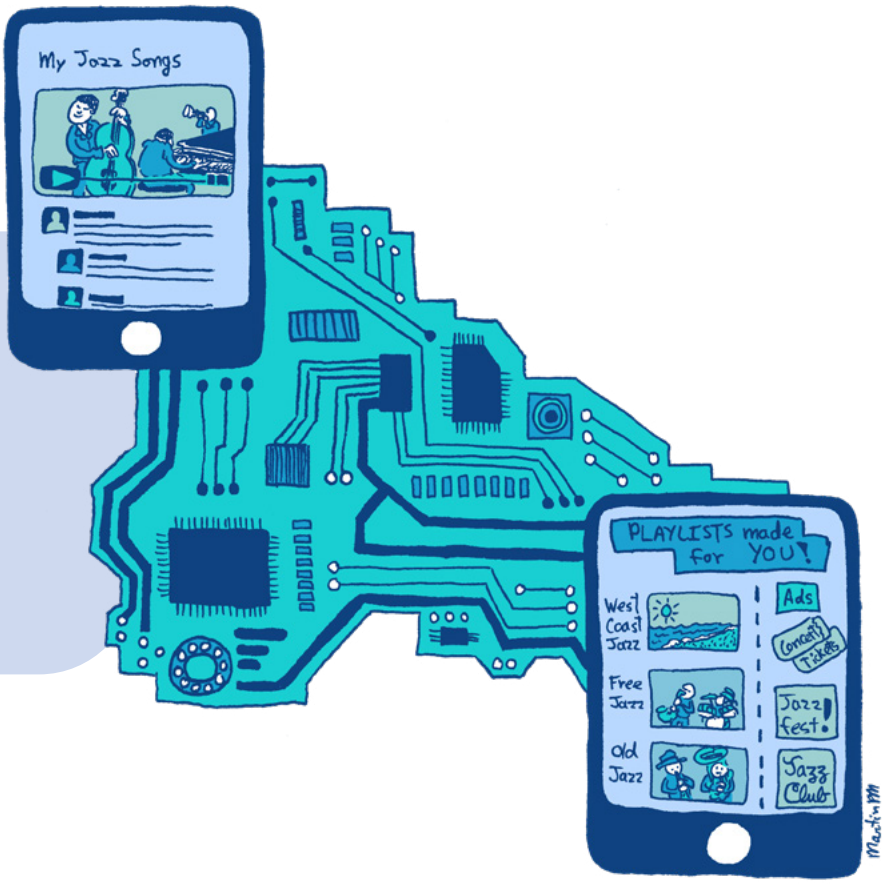
AI is the set of computer techniques that enable a machine (e.g. a computer or telephone) to perform tasks that typically require intelligence, such as reasoning or learning. It is also referred to as the automation of intelligent tasks. Scientific developments in AI, such as deep-learning techniques, have made it possible to design high-performance intelligent devices, with access to huge amounts of data and ever-increasing computing power. These new techniques have been rapidly deployed on a large scale in all areas of social life, in transport, education, culture and health.

WHAT DOES AI DO?

AI is based on the use of algorithms that process data. An algorithm is a sequence of instructions that can be used to solve problems and accomplish complex tasks. This series of steps transforms input information into a useful result (output). A recipe is a kind of algorithm: to cook a dish (output), you need to have the right ingredients (input) and follow instructions on how to use them correctly (the algorithm). The sequence of instructions that a computer uses to predict a person's age (output) from their picture (input) is also defined by an algorithm.

Algorithms developed in the area of AI are used to perform functions.

Here are a few examples.



PREDICTION

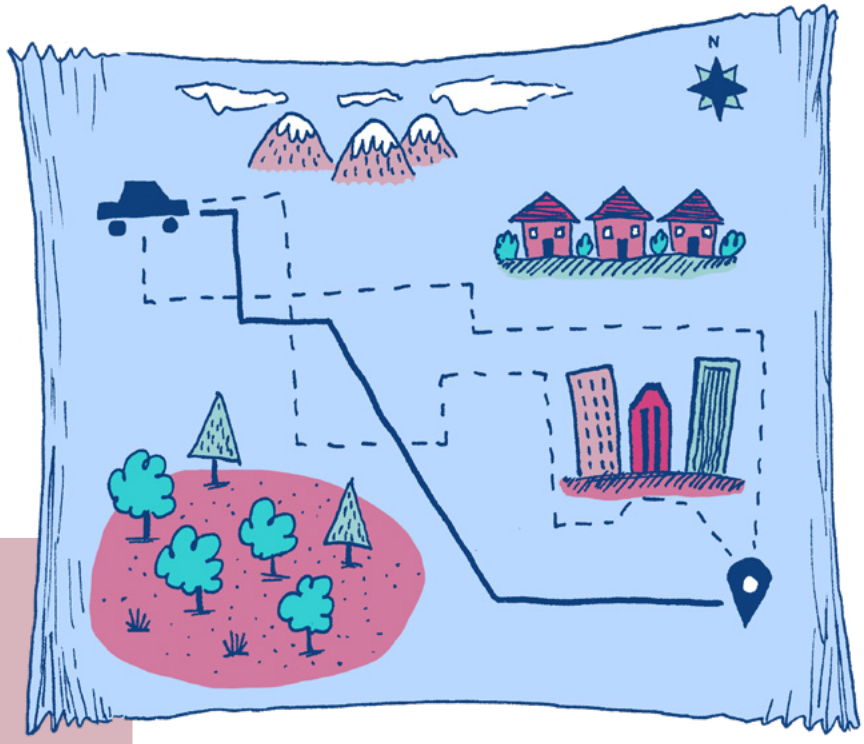
(e.g. predicting an internet user's interest in a type of cultural content based on their browsing history)

DETECTION

(e.g. detecting if and where a face appears on an image)

IDENTIFICATION

(e.g. finding a person's name from their photo)

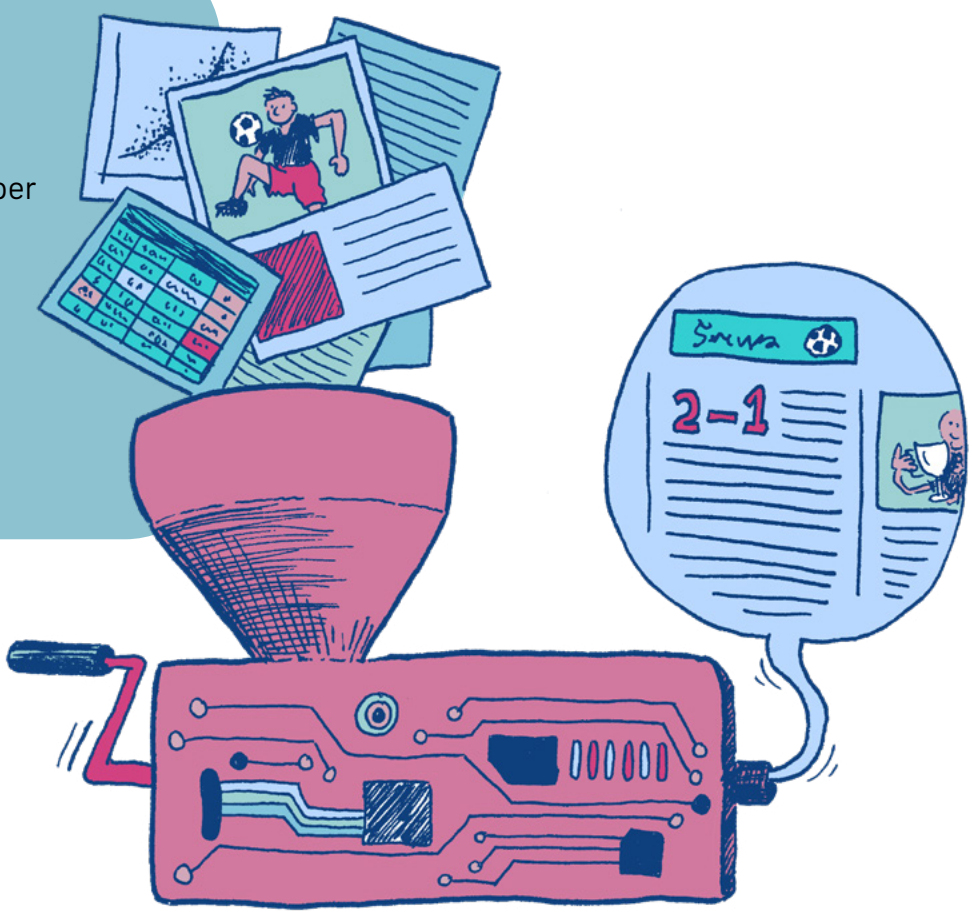


PLANNING AND DECISION-MAKING

(e.g. choosing the fastest route to the hospital based on traffic information)

CONTENT GENERATION

(e.g. generating a newspaper article about the score of a sporting event)



The combination of these functions allows for several high-level tasks to be performed.

Here are a few examples.

> PERCEPTION

When AI is used to analyze measurements, such as a camera image or microphone recording. For example, an algorithm can detect spoken or handwritten words, or determine whether there is a gathering of people from a photograph.

> PROCESSING OF NATURAL LANGUAGE

This is when AI is used to process the meaning of words. It can understand a command, such as calling a contact, or predict the next word to be written, such as on smartphone keyboards. It is also possible to detect whether a news article will be of interest to any given person or to generate a translation of a paragraph in a foreign language.

> ROBOTICS

This is when AI controls a machine that can act in the physical world. Based on the information received by its sensors, AI must plan and make the best decision in order to accomplish its objective. For example, a self-driving vehicle must drive safely to a given destination, and a robot surgeon must best replicate the movements of the human controlling it.

> OPTIMIZATION AND PROBLEM SOLVING

This is when AI has to resolve a situation in a defined environment. For example, making decisions in a video game, or planning the optimal route for car or taxi trips.

A wide variety of **application domains** can benefit from the ability to automate these functions, including education, journalism, cybersecurity, video games, art (music, cinema, etc.), finance, health care, transportation, military, ecology, climate science, etc.

TO TAKE IT FURTHER

There are **two major technical areas** in AI. Algorithm developers can either give all the **instructions** to the machine in advance, or allow the machine to **learn** the steps by itself. Writing the rules in advance requires that the developers themselves know **how to solve** the problem they are asking AI to solve. In some cases, this is possible. For example, scientific models based on the laws of physics are used to predict the weather. However, when the problem is too **complex**, AI developers prefer to choose machine learning. In this case, the model has to be **taught**. For example, it is very difficult to write the rules that allow a computer to differentiate a cat from a dog in a picture, as these two species vary greatly, and moreover, pictures can be taken from different angles, under different lighting, etc. Machine learning algorithms solve a problem by showing several images of dogs and cats to a computer and teaching it to differentiate the animals by itself. Of course, the performance of the algorithm will depend on the **quantity** of images, as well as their **representativeness**—if only black cats seen from the front have been shown, the algorithm will not know what to do with a picture of a white cat taken from above. That's why AI algorithms usually require a large amount of data, and it matters how this data is collected.

WHAT AI DOES NOT DO

At the present time, there is no "general" or "strong" AI that, like human intelligence, can perform various tasks such as play chess, drive a vehicle or recognize a tumour, for example. This is a goal that some researchers in this area strive to achieve, but the most advanced systems that currently exist remain far from this point. Today's AI is described as "weak" because, though it might perform certain tasks more efficiently than a human, it can only do specific tasks for which it was developed. Some people think that general AI may one day express emotions or self-awareness. We are not there yet, and all that "weak" AI can do is identify emotions and simulate them.

ESSENTIAL CONCEPTS

INTERNET OF THINGS
This refers to an infrastructure of interconnected objects capable of communicating with each other without any human intervention.

ALGORITHM
This means a sequence of instructions that transforms an input into an output. For example, a pancake recipe allows you to turn ingredients into a delicious meal by following specific steps. The steps to solve a Rubik's cube are also an algorithm.

MASSIVE DATA, MEGADATA OR BIG DATA
These are datasets so large that they cannot be collected, stored and analyzed using conventional methods. Many AI algorithms use big data.

MACHINE LEARNING
This refers to the ability of a machine to learn how to perform a task without instructions but rather through experience acquired during a practice process.

DEEP LEARNING
This refers to machine learning with a specific and particularly powerful technique, which uses neural networks connected on several layers. These artificial neural networks, inspired by the functioning of neurons in the brain, consist of mathematical processing of incoming data.

Part 2

Ethical Artificial Intelligence

While AI technologies may be promising, their use raises ethical and social concerns on which we need to reflect collectively, considering their growing influence on society. This is one of the objectives of ethical AI, which strives to identify and prevent the misuse of AI while promoting its benefits.

WHAT IS ETHICAL AI?

Ethics refers to a set of standards (principles and rules) that we must abide by if we want to do the right thing, such as the principle that we must not harm others or the rule that forbids lying. Ethics is said to be prescriptive because it prescribes what should be, what needs to be or what is acceptable according to the values that are embraced. Ethical standards articulate values that we recognize as moral. Ethics also refers to the philosophical discipline that attempts to identify these standards and values.

Ethical AI is the set of standards and values applied to the development and use of AI. Thus, it is a restricted area of application of ethics, but since AI technologies are disrupting social organization and can have very profound harmful consequences, this area of ethics appears to be crucial and is undergoing major development.

Lastly, ethical AI is part of public ethics (i.e. ethics applied to societal controversies that require a solution acceptable to the individuals who disagree with it). This is also the case with environmental ethics or bioethics.

SOME ETHICAL AND SOCIETAL CHALLENGES

EExperts, researchers and citizens have already expressed many concerns about the development and use of AI. These include:

> PRIVACY RISKS

Privacy is a social value that has recently emerged in societies that aspired to democracy. It is now widely accepted and many people are concerned about the violation of their privacy. The risk of invasion of privacy is increased by the unprecedented performance of increasingly intrusive means of data collection and by the multiplication of places where personal data is collected (e.g. mobile phones or connected devices) in the home. This risk is also increased by the new analytical possibilities offered by AI. For example, algorithms can identify individuals by cross-referencing different inputs that were made anonymous.

> THE RISK TO LIMIT **FREEDOM OF CHOICE AND INDEPENDENCE**

Freedom of choice and independence (i.e. the ability to make decisions) are generally valued. But machines can severely restrict our freedom by making decisions for us without our awareness or ability to challenge them. One example is the "bubble filtering" effects of algorithms that offer users content that is ever consistent with their digital behaviour (i.e. their previous choices), thus limiting the diversity of content offered to them or the chances of discovering new preferences. This is the case when music content sites propose songs that remain similar to what the user is listening to.

> THE **RISK OF DISCRIMINATION**

Discrimination is treating similar cases differently without good reason. It is the opposite of justice, which is to treat similar cases the same way. A smart machine can reproduce or generate discrimination when its algorithm or the data it learns from contains errors or biases that lead to individuals or groups being treated differently from the rest of the population without acceptable justification. As a result, individuals or groups are excluded from the benefits of AI. A machine that is populated with data that does not cover the entire population will deprive part of the population of its benefits; this is the case when a machine that detects skin cancer is taught to detect it only on light skin. It will not detect with the same reliability the cancers that appear on dark skin.

> LACK OF **EXPLAINABILITY**

It is often difficult to explain how an algorithm reached a specific decision or recommendation. Smart devices are like "black boxes" that we do not understand. However, it seems important to be able to understand a decision made by an algorithm in order to be able to keep control of the decisions that affect us, and to be able to challenge or even change them. "Explainability" is a value closely linked to freedom of choice and independence but also to justice, such as when smart devices reproduce or generate discrimination.

> AUTOMATION OF WORK

While smart technology makes it possible to perform repetitive tasks and can thus reduce the drudgery of work, it is replacing human workers. In addition to the social and economic consequences of this replacement, such as increasing unemployment, it impacts the sense of solidarity and self-respect. One of the ethical issues at hand is maintaining human interaction, particularly in sectors such as health and education; another is preserving respect and self-esteem, which are based, in part, on a sense of social usefulness.

> **ENVIRONMENTAL** RISKS

Environmental protection and the fight against climate change are now major societal challenges. Deploying AI can help improve our collective and individual actions to reduce our negative impact on the environment, but if left unchallenged, it can also increase this negative impact. Addressing the environmental issue surrounding AI implies taking into account the impact of all the elements that make the use of AI possible, such as smart phones with their polluting components or data centres. For example, mega data storage centres (data centres), which in particular enable machine learning, are major energy consumers.

PRINCIPLES IN ACTION

In an effort to provide socially acceptable responses to the various ethical, societal and political concerns raised by the deployment of AI, numerous declarations of ethical principles and guidelines have been produced around the world. These documents attempt to define the ethical principles that will guide reflection in order to limit the negative consequences of AI use. This is the case, for example, with the principles adopted by the Organisation for Economic Co-operation and Development (OECD) and the G20 countries, the European Commission's Ethics Guidelines for Trustworthy Artificial Intelligence, or the Montreal Declaration for a Responsible Development of Artificial Intelligence, which is distinguished by a deliberative process that has informed the work of experts. This declaration promotes 10 ethical principles for a responsible development of AI, such as the principle of well-being:

The development and use of artificial intelligence systems (AIS) must permit the growth of the well-being of all sentient beings.

The principles are intended to guide action and can thus form the basis for various concrete measures in the management and ethical development of AI. These may include, for example:

> legislation;
> public policies;
> audits or certifications;
> training;
> institutional stakeholders;
> codes of conduct;
> technical solutions.

ESSENTIAL CONCEPTS

ETHICAL DILEMMA
A situation poses an ethical dilemma when it requires you to choose between two incompatible and both morally unsatisfactory options. This happens when there is a conflict of principles or values and a significant moral value must be sacrificed.

ALGORITHMIC GOVERNANCE
This may refer either to how algorithms are controlled, how they are developed and how their use is monitored, or to how algorithms control or regulate our lives, our social relationships and our public institutions.

BLACK BOX
A black box is an AI system by which it is difficult or impossible to explain decisions or recommendations being made. This expression is used to emphasize the lack of transparency in the operation of smart devices, especially those based on deep learning algorithms.

ETHICS BY DESIGN
One way of regulating the ethical use of digital tools and AI systems is to consider ethical principles when they are being designed by researchers and engineers. This prevents their unethical or socially undesirable uses from the start.

BIAS
This refers to the thinking process that alters and distorts judgment. An algorithm may be biased if it favours certain outcomes without a good moral justification.

THE MONTRÉAL DECLARATION FOR A RESPONSIBLE DEVELOPMENT OF AI

The Montréal Declaration for a Responsible Development of AI (the Declaration) is a coherent list of ten ethical principles that aims to provide moral guidance for thinking about responsible and socially acceptable artificial intelligence, and to frame its development, deployment and use in various sectors of social life.

The Declaration is the result of an inclusive deliberative process that brought together citizens, experts, public officials, industry stakeholders, civil society organizations and professional associations. This University of Montreal initiative brought together over 500 participants in deliberative workshops.

The interest of this deliberative process is threefold:

1. To collectively arbitrate ethical and societal controversies on AI.
2. To improve the quality of reflection on responsible AI.
3. Strengthen the legitimacy of proposals for responsible AI.

The principles are not hierarchical. The last principle is not less important than the first. It is possible, depending on the circumstances, to give more weight to one principle than to another, or to consider that one principle is more relevant than another. This allows for a legitimate diversity of interpretations, but it does not allow for just any interpretation.

It should also be emphasized that these are ethical principles, not rules of governance, let alone legal norms. However, they can be translated into political language and interpreted in a legal manner, notably in the language of fundamental human rights.

Finally, although these principles were developed in a given society, Quebec and Canadian society, they provide a basis for intercultural and international dialogue.

The Declaration is the most comprehensive AI ethics document to date and remains open to discussion and revision.
Here is the list of the ten principles which are declined in about sixty sub-principles (to be discovered on the site: declarationmontreal-iaresponsable.com):

1. WELL-BEING PRINCIPLE

The development and use of artificial intelligence systems (AIS) must permit the growth of the well-being of all sentient beings.

2. RESPECT FOR AUTONOMY PRINCIPLE

AIS must be developed and used while respecting people's autonomy, and with the goal of increasing people's control over their lives and their surroundings.

3. PROTECTION OF PRIVACY AND INTIMACY PRINCIPLE

Privacy and intimacy must be protected from AIS intrusion and data acquisition and archiving systems (DAAS).

4. SOLIDARITY PRINCIPLE

The development of AIS must be compatible with maintaining the bonds of solidarity among people and generations.

5. DEMOCRATIC PARTICIPATION PRINCIPLE

AIS must meet intelligibility, justifiability, and accessibility criteria, and must be subjected to democratic scrutiny, debate, and control.

6. EQUITY PRINCIPLE

The development and use of AIS must contribute to the creation of a just and equitable society.

7. DIVERSITY INCLUSION PRINCIPLE

The development and use of AIS must be compatible with maintaining social and cultural diversity and must not restrict the scope of lifestyle choices or personal experiences.

8. PRUDENCE PRINCIPLE

Every person involved in AI development must exercise caution by anticipating, as far as possible, the adverse consequences of AIS use and by taking the appropriate measures to avoid them.

9. RESPONSIBILITY PRINCIPLE

The development and use of AIS must not contribute to lessening the responsibility of human beings when decisions must be made.

10. SUSTAINABLE DEVELOPMENT PRINCIPLE

The development and use of AIS must be carried out so as to ensure a strong environmental sustainability of the planet.

Part 3
Deliberating on ethical AI

ENGAGING CITIZENS

There are many ways to engage citizens in the ethical AI debate. Depending on the level of engagement, three typical processes can generally be distinguished:

> CONSULT

Consultation is a process which consists of gathering opinions already formed by those consulted on a previously defined topic. Consultation allows the individuals being consulted to ask questions, and to express their concerns, expectations, comments or opinions in order to improve decision-making.

> DELIBERATE

Deliberation is a rational discussion through an exchange of arguments for a collective decision. Deliberation should increase the knowledge of each participant and allow for a better understanding of individual and collective interests. It can alter our initial preferences. It does not necessarily lead to consensus, but rather to the identification of common orientations based on convergences and divergences of opinion and the reasons behind them.

> CO-CONSTRUCT

To co-construct means to engage citizens during the entire process of ideation and creation. It is a collaborative and interactive process by which citizens and stakeholders exchange and create together.

WHY SHOULD WE DELIBERATE ON ETHICAL AI?

The deployment of AI affects all spheres of personal and social life. It affects everyone, and no one can measure all the implications of a complex technological and social phenomenon. It is crucial to expand expertise: that of scientists, of course, but also of citizens, AI users and those affected by it. This is why it is essential to use **collective intelligence** and to involve as many people as possible, beyond the circles of experts and public decision makers, in the process of reflecting on the social and ethical issues surrounding AI.

Deliberation not only deepens our knowledge about AI as a technological object that transforms our social and political relationships, it also allows us to make more informed decisions and gives these decisions a sense of legitimacy that is often missing from those made by experts. This requires for a large number of individuals and a wide variety of participants to be engaged. The cultural and social wealth in the world is the only limit.

Finally, by engaging in deliberations, we can **have our voices heard** individually and collectively, and have the opportunity to use AI for the common good and for our fundamental interests.

ESSENTIAL CONCEPTS

ARGUMENT
An argument is a form of reasoning that proves or justifies an assertion (an opinion). To defend an opinion and convince the persons we are speaking to with good reasons, we need to use a coherent set of arguments.

COMMON GOOD
The common good refers to a reality shared by all, regardless of the social organization. "Common" implies the idea of a link between members of a group. An initiative for the common good implies that it is developed in the interest of all.

COLLECTIVE INTELLIGENCE
Collective intelligence is the ability of a group to come up with more appropriate solutions, make better decisions, and increase our knowledge by discussing, exchanging arguments, and sharing individual knowledge. Practising collective intelligence requires that group members share common goals and interests, as well as a collaborative space (physical or virtual).


Deliberations bring together a diverse set of people with varying opinions on a given topic.
It entails thinking collectively by exchanging and bringing forth arguments.
Common ground? Compromise?
The goal is to think collectively toward a consensus.
Consensus doesn't mean that everyone agrees. Consensus is a position or an opinion that seems most reasonable for the community.
Consensus
MartinPM

TO DELIBERATE EFFICIENTLY, LET'S REMEMBER THAT...

... participants are equal in discussions, and participation in a deliberative workshop requires mutual respect.

... all opinions matter: opinions expressed in good faith should not be excluded without discussion.

... the opinions expressed must be supported by arguments and the exchange of arguments must be public.

ACKNOWLEDGEMENTS AND CREDITS

EDITORS

DILHAC, Marc-Antoine
Université de Montréal, OBVIA, Mila-Institut Québécois d'intelligence artificielle, Chaire CIFAR, Algora Lab.

MAI, Vincent
Université de Montréal, Mila-Institut Québécois d'intelligence artificielle, Algora Lab.

MÖRCH, Carl-Maria
Université de Montréal, OBVIA, Algora Lab.

NOISEAU, Pauline
Université de Montréal, OBVIA, Algora Lab.

VOARINO, Nathalie
Université de Montréal, OBVIA, Algora Lab.

ILLUSTRATOR

PATENAUDE-MONETTE, Martin
martinpm.info

GRAPHIC DESIGNER

HAUSCHILD, Stéphanie
stephaniehauschild.com

TRANSLATION

CESARIO, Bianca
Révidaction

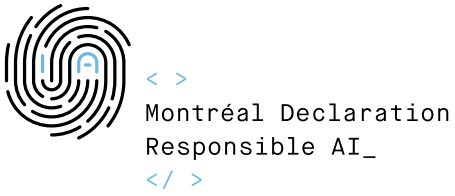
WITH THE CONTRIBUTION OF

FLORES ECHAIZ, Lucia
Université de Montréal, OBVIA, Algora Lab

LANTEIGNE, Camylle
Université de Montréal, OBVIA, Algora Lab

SALAZAR GOMEZ, Fatima Gabriela
Université de Montréal, OBVIA, Algora Lab

PARTNER INSTITUTIONS



Thanks to the financial support of the Government of Quebec, the Government of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche du Québec and the National Research Council of Canada.