



&lt; &gt;

# Montréal Declaration Responsible AI\_

&lt;/ &gt;

## PART 2

# 2018 OVERVIEW OF INTERNATIONAL RECOMMENDATIONS FOR AI ETHICS



This document is part of the 2018  
**MONTREAL DECLARATION FOR  
A RESPONSIBLE DEVELOPMENT  
OF ARTIFICIAL INTELLIGENCE.**  
You can find the complete report [HERE](#).

## WRITTEN BY

**MARTIN GIBERT**, Ethics Counsellor at IVADO  
and researcher in Centre de recherche  
en éthique (CRÉ)

**CHRISTOPHE MONDIN**, Research  
professional for CIRANO

**GUILLAUME CHICOISNE**, Scientific Programs  
Director, IVADO

## TABLE OF CONTENTS

<b>1. INTRODUCTION</b>	<b>80</b>
1.1 Methodology	80
1.2 Opening remarks	83
<b>2. THEMATIC SUMMARY OF RECOMMENDATIONS</b>	<b>84</b>
<b>3. REPORTS ON AI DEVELOPMENT: TECHNICAL DATA SHEETS</b>	<b>92</b>
3.1 The seven reports studied	92
3.2 Reports examined, but not selected	94
3.3 Other reports consulted	97
CREDITS	I
PARTNERS	II
<b>TABLES AND FIGURES</b>	
Table 1: Occurrence of key concepts in the seven documents examined	81

# 1. INTRODUCTION

In December 2016, Corinne Cath and her colleagues from Oxford University and the Alan Turing Institute published a comparative analysis of artificial intelligence policies from the European Parliament, the House of Commons of the United Kingdom, and the White House<sup>1</sup>. They concluded that these three reports correctly identified various ethical, social and economic issues, but lacked a long-term strategy to develop “good AI”. Where do things stand today? How do various government and non-government organizations foresee the changes that AI will bring to society?

Keep in mind that many events have occurred since December 2016 which have changed public and government expectations of AI, and information technologies in general. The first self-driving car crashes have occurred. Revelations on the attempted tampering with the latest American presidential elections via Facebook as well as the Cambridge Analytica scandal which blew up in March 2018, elicited strong reactions and sparked fear for the good health of democracies. Likewise, Google’s image has been somewhat tarnished from its veiled collaborations with the American army. We will have a more accurate understanding of the reports analyzed in this document if we put them back into context—this is especially true for the declaration of ethical principles published by Google in June 2018.

## 1.1

### METHODOLOGY

To provide a brief overview of the situation in 2018, we have analyzed seven recently published reports and declarations of principles. The technical sheets on the selected documents are detailed in the third section of this document. We have added files from reports that were examined, but not selected. What initially guided our choice were ethical recommendations, but that is far from always being the case. In fact, much prospective thinking on the future of AI is from a chiefly economic perspective: how, for example, can we develop an ecosystem that fosters innovative AI companies, what is the strategic plan for AI development in a given country? We set aside reports, therefore, that were primarily economic as well as economic recommendations in the reports selected. Moreover, we did not select reports that focussed exclusively on one specific field, such as robotics research ethics or self-driving car regulations. The goal was to examine a general set of recommendations that could be compared to one another.

We also sought a certain diversity when making our selection, to give us a broad enough scope for comparison. That is why two of the reports (Villani and CNIL) are in French, and the other five are in English. One report is from a private company (Google), three are from non-governmental organizations (IEEE, Asilomar and AI Now) and three others present the official policies of several countries (UKRS, Villani and CNIL). Some reports, therefore, are more global in vision, whereas others are more local. Moreover, some reports were relatively concise (Asilomar, Google, AI Now), while others were much longer and detailed, particularly because they included economic considerations.

In the technical sheets in section 3, we also highlighted clearly identifiable principles and recommendations. We call “principles” the very general proposals, such as “AI should be beneficial for society”, whereas the “recommendations” are more targeted and relatively concrete, such as “we must develop standards to track the source and use of data sets throughout their entire life cycle”.

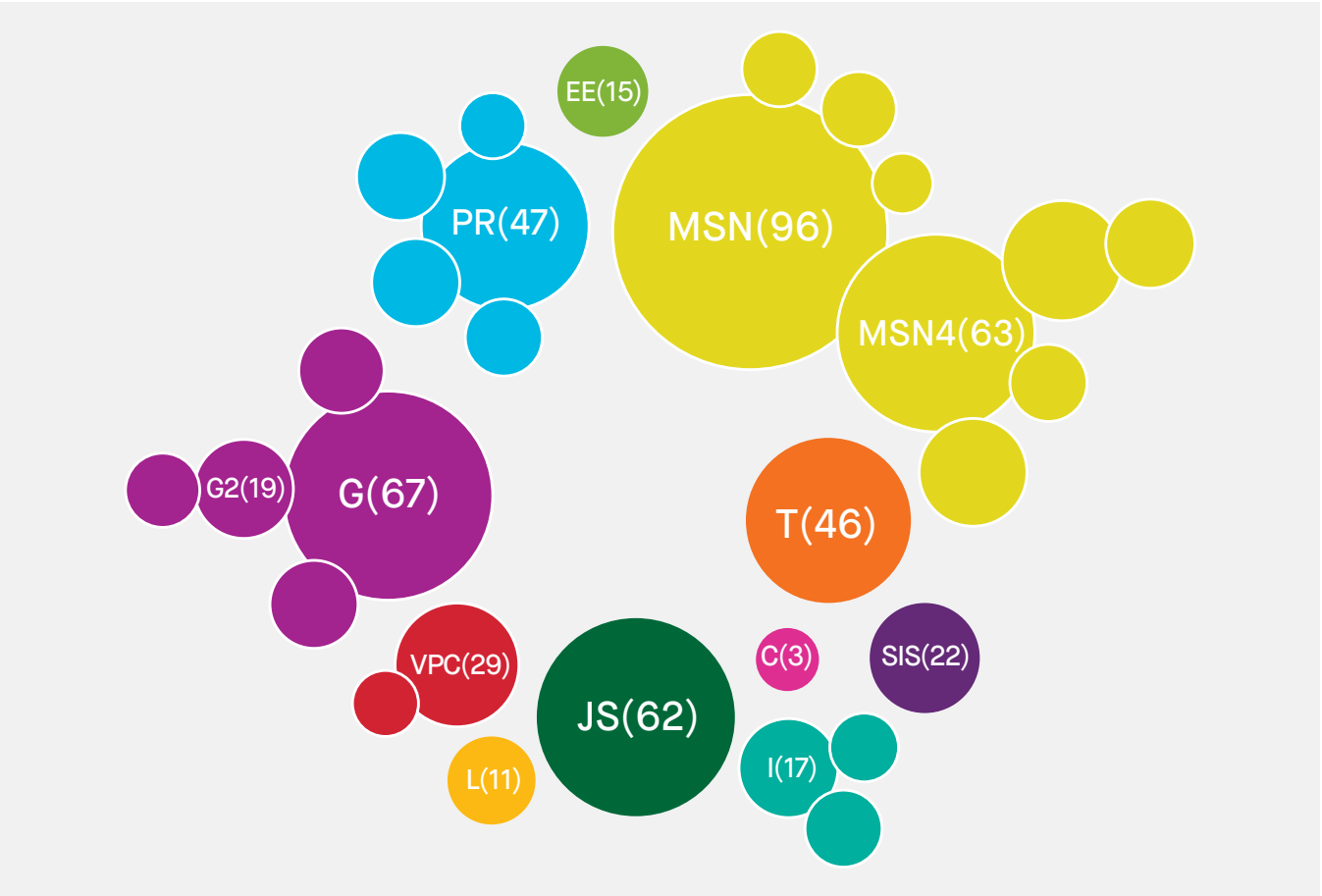
<sup>1</sup> Cath, C., Wachter, S., Mittelstadt, B. et al. Sci Eng Ethics (2018) 24: 505. <https://doi.org/10.1007/s11948-017-9901-7>

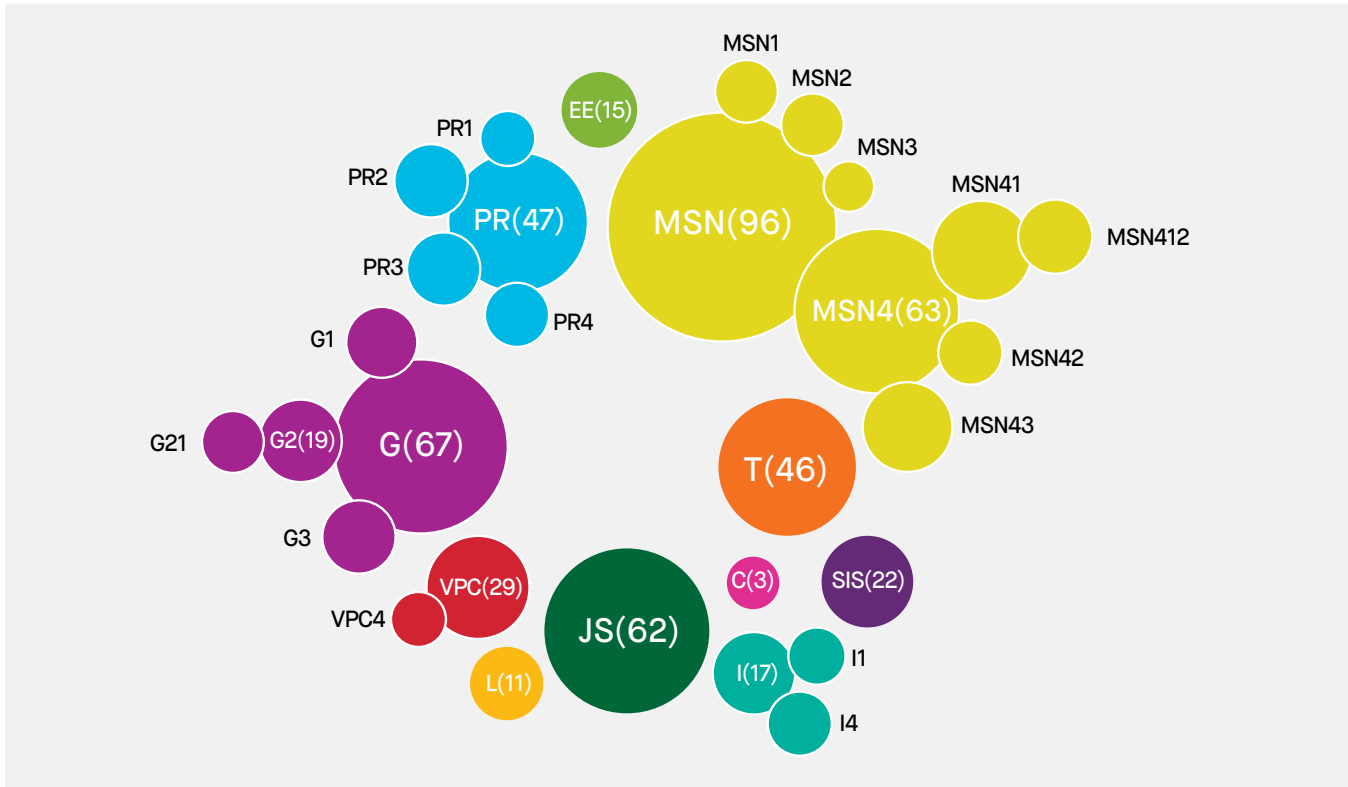
From a methodological standpoint, we started by identifying the ethical recommendations in the seven reports. We retained 230 recommendations. We then classified these recommendations into one of seven categories taken from the preliminary version of the Montréal Declaration for a Responsible Development of Artificial Intelligence: well-being, autonomy, justice, privacy, knowledge, democracy and responsibility—one recommendation may apply to many categories. The advantage of these labels is that they directly refer to what interests us, namely moral values. Of course, classifying a recommendation is often a matter of interpretation

and other analysts may have reached different conclusions. We then summed up each value, and presented the results in the second section.

In order to shed new light on the recommendations, we also categorized them according to a set of well-defined key concepts. These concepts are taken from an index developed from citizen recommendations established during the collective reflection (coconstruction) sessions on the Montréal Declaration. This is how we obtained the graphic below:

Table 1 : Occurrence of key concepts in the seven documents examined





## LEGEND

C	Consent
EE	Environment and ecology
G	Governance
G1	Collectivism/Individualism
G2	Democratic governance
G21	Digital commons
G22	Citizen participation
G3	Public/private governance
G31	Conflicts of interest
G32	Public institutions/Private companies
G33	Monopoly
I	Influences
I1	Lobbyism
I2	Manipulation
I3	Paternalism
I4	Vulnerability of people
JS	Social justice
L	Freedoms
MSN	Socio-digital mutations
MSN1	Acceptability
MSN2	Activity transformations
MSN3	Respecting humans
MSN4	AI skills
MSN41	Human skills

## LEGEND

MSN411	Dependence on technology
MSN412	Digital literacy
MSN413	Transformation of human skills
MSN42	Human-AI synergy
MSN43	AI skills
PR	Sharing responsibility
PR1	Disempowerment
PR2	Accountability
PR3	Shared responsibility
PR4	Decision sovereignty
SAA	Stress, alarmism and anxiety
SIS	Safety and system integrity
T	Transparency
VPC	Privacy and confidentiality
VPC1	Anonymity
VPC2	Confidentiality
VPC3	Right to be forgotten
VPC4	Data ownership
VPC5	Intrusion

## 1.2

### OPENING REMARKS

Before presenting the reports and different summaries per value, we felt a few general remarks were in order. First, the similarities among the reports can be striking: it is often difficult to detect any major divergence among the recommendations from the seven reports. This can partly be explained by research consensus: these reports seek to bring people together, not stir controversy, and they avoid potentially divisive subjects by remaining quite general overall. But it could also be that this convergence simply reflects a fundamental agreement on the types of relationships that we should maintain with AI as a whole. After all, it is hardly surprising that everyone agrees to fight discrimination caused by algorithmic automation, or to promote reinforcing consent when managing user data.

These similarities may also be explained by the fairly homogenous character of the societies these reports are from: rich occidental countries that globally share the same democratic and liberal values. We need, therefore, to address the elephant in the room: how do we regulate AI on an international scale? Data, information and algorithms seem especially impervious to territorial boundaries. What authorities in the United Kingdom, France or any other country can accomplish will always remain very limited, then, in the absence of international cooperation. But is it truly feasible? We must also not forget that calls to reduce discrimination and increase equality exist within a global context of growing inequalities. In other words, it is difficult to isolate issues of AI ethics from issues of international justice.

While similarities exist in the reports we examined, they still contain what could be considered different areas of focus. Some reports highlight political and economic issues (Villani and UKRS) while others concentrate on legal or ethical considerations. Moreover, though they are all presented as reports from experts, the report by CNIL is based, in part, on citizen consultations. The declaration of principles by Google is unique in that it is the only private company

represented among these reports. Its declaration could create a potential conflict of interest, but it also is the most likely to have a tangible international impact, given the power of the company.

In terms of content, the most striking difference is in the self-regulation of companies and the role of public bodies in AI system governance. It comes as no great surprise, then, that reports issued by the government, such as the Royal British Society, the “UKRS”, or the “Villani report” commissioned by the French government offer more potential solutions from public institutions. They also largely favour legislative tools to meet the challenges the arrival of AI systems heralds—this is also the point of view held by the Institute of Electrical and Electronics Engineers (IEEE). On the contrary, the AI Now and Asilomar reports broach the issue from the perspective of companies that can develop safety tools, self-regulation rules and best practices guides. The CNIL report stands out by suggesting two new principles—vigilance and loyalty of AI systems—while the Villani report pays considerable attention to ecological issues.

Lastly, the pragmatic or prosaic language of these reports is worth mentioning. We are far from the lyricism and existential considerations found in the works of Yuval Harari, Nick Bostrom or sci-fi literature. The focus is not placed on the radical shift that AI is creating in human history, but on a cautious and progressive adaptation of technological innovations. Seen this way, it is worth reiterating the conclusion that Corinne Cath and her colleagues arrived at after reading through the 2016 reports: the general and long-term vision of society with “good AI” is still a work in progress.

## 2. THEMATIC SUMMARY OF RECOMMENDATIONS

The seven reports or documents quoted in the next section are:

- > **AI Now:** the 2017 report AI Now Institute.
- > **Asilomar:** the principles that emerged during a Future of Life Institute Conference.
- > **CNIL:** the report from the Commission nationale (française) de l'informatique et des libertés.
- > **Google:** the principles published by Google in June 2018
- > **IEEE:** the report from the Institute of Electrical and Electronics Engineers
- > **UKRS:** the report from the British Royal Society
- > **Villani:** the report "Donner un sens à l'intelligence artificielle" led by French MP Cédric Villani

### WELL-BEING

Every report that we examined contained recommendations explicitly associated with well-being. They appear the most often, which is unsurprising, given that this value is key, and even at times synonymous, with the concept of good. The recommendations associated with well-being are particularly associated with the values of AI skills, social justice, safety and system integrity, privacy and confidentiality, human-AI synergy, and collectivism/individualism.

We note that certain trends start to emerge in the reports. AI Now highlights the challenges of discrimination and biases by demanding, for example, that AI systems that impact society as a whole be developed by people that represent society in all its diversity (AI Now, p. 2). (Villani goes a step further by specifying that every level of the AI design chain

must be representative of society [Villani, p. 23].) For its part, CNIL focuses on algorithm loyalty towards people so as to not "betray" them by reinforcing discrimination (CNIL, p. 48). The IEEE puts safety first (IEEE, p. 22) for AI systems, which should always be designed to benefit humans.

Asilomar views it from a research perspective: the goal should not be to create neutral intelligence, but beneficial intelligence (Asilomar); that is why funding should be allocated to this end (Asilomar) and include disciplines such as social sciences, ethics, law, public health or ecology (Asilomar). This is also the case for the UKRS, which demands that the government foster research by developing data sharing standards (UKRS, p. 8) and educate machine learning developers on social and ethical issues (UKRS, pp. 9 and 12). The UKRS also stands out by its focus on research and teaching.

Villani pays special attention to the effects of workplace automation as well as many economic considerations. He recommends, for example, that we create a "public lab for labor transformations" and "launching a legislative reform" (Villani, p. 12) of working conditions in the age of automation. These recommendations fall within a larger project that underscores the general interest and issues of common good, particularly health care: we must develop AI to ensure the "early detection of diseases, the 4 Ps of healthcare [predictive, preventive, personalized and participative], the elimination of medical deserts, emission-free urban transport" (Villani, p. 9). The Villani report is also the only to mention how to promote the ecological transition (Villani, p. 14), which has an obvious impact on well-being.

Google, finally, addresses the notion of well-being in its first principle by stating that AI should be socially beneficial. The principles of the company differ from other reports in that they focus on doing no harm over promoting well-being: it is important, then, to conduct tests to "avoid unjust impacts", limit prejudicial or abusive uses, and not develop potentially destructive technology.

But whose well-being is actually being discussed in these reports? The focus, more or less explicitly, is always on the well-being of humans: IEEE maintains,



for example, that human well-being must be made a priority, using the best available and generally accepted indicators of well-being as a reference point (IEEE, p. 25). No report mentions the well-being of animals. Likewise, when ecological issues are raised (Villani, p. 14) it is from an anthropocentric perspective (as opposed to a pathocentric, biocentric or ecocentric perspective). This does not mean that non-human well-being is not worthy of discussion. In fact, the idea of aligning AI with human values, which is found in Asilomar, leaves the door open to extending compassion towards those who are most vulnerable or concern for other species as human values.

Though they address only human well-being, the reports are “universalist” in that they make no distinction between subcategories of the human population—in other words, it is a question of respecting the universality of human rights. For example, no report claims that only an oligarchy, a state or an organization should benefit from AI—quite the opposite, specifies Asilomar. In other words, as Villani insists, the opportunities associated with the arrival of AI must benefit everyone (Villani, p. 23). He also notes that we must anticipate the impact of technological changes, “which often hit the most fragile portions of the population the hardest” (Villani, p. 14).

When the subject of wealth created by AI (a question that political philosophers call distributive justice) is broached, the reports are careful not to speculate on who should benefit from it. They particularly call for reflection on the matter. Villani recommends “initiating dialogue with industrial partners on how value-added is shared” (Villani, p. 13) while the UKRS advocates that society urgently consider the way “the benefits of automated learning can be distributed among society” (UKRS, p. 12). This “time to reflect” on wealth redistribution is echoed in the fairly common plea in all reports to enhance AI research through collaborations with social sciences or ethics (e.g. Asilomar).

For its part, the preliminary version of the Montréal Declaration suggests the following principle: “AI development should ultimately aim for the well-being of all sentient beings.” It takes a more inclusive stance by adopting a pathocentrist view. That is

perhaps one of the most original elements of the Montréal Declaration: to consider not only the fate of human beings, but of all individuals that could be affected by AI development.

## AUTONOMY

Recommendations explicitly tied to the notion of autonomy are present in every report—with the exception of AI Now. These recommendations are closely linked to issues of human skills, human-AI synergy, AI skills, acceptability, vulnerability of people and social justice.

Overall, the idea that AI must respect human autonomy is defended throughout the different reports. Asilomar states, for example, that AI systems must be designed and operated so they are compatible with the ideals of human dignity, respect of rights, freedoms and cultural diversity. The CNIL (CNIL, p. 57) takes it perhaps one step further: not merely respect autonomy but promote it, starting at the design phase. Among philosophers, this distinction between respecting and promoting generally refers to choosing between a deontological logic of respecting standards (autonomy as a right) and a consequential logic of promoting values (autonomy as a good). However, we must avoid over-interpreting the choice of terms. The CNIL even specifies that it is a matter of correcting a situation since it insists on the importance of “overcoming asymmetries”, given that there can be no true autonomy in a situation where one stakeholder holds all the power or all the information. For the CNIL, promoting autonomy is also a question of raising awareness among professionals who use AI (CNIL, p. 55).

Respecting or promoting user autonomy is also expressed in the idea that AI must remain a tool, an instrument that serves users or, broadly speaking, human beings. The IEEE notes that AI systems should always be subordinate to human judgment and control (IEEE, p. 23). This idea echoes the Google principle that AI technologies “must be subjected to appropriate human direction and control” (Google). The CNIL report, incidentally, is named “Comment



permettre à l'homme [sic] de garder la main" (How can humans keep the upper hand?).

This quest for autonomy could be the result of a joint effort between businesses that offer AI and those who use it. For Asilomar, human beings must decide if and how to delegate decision-making to AI systems so they can accomplish goals determined by humans. The CNIL (CNIL, p. 57) offers more concrete recommendations by noting that users should be able to "play with" the parameters of a given system, which has the advantage of fostering understanding. For Google, information and consent must guide how companies use AI, especially "by providing appropriate transparency and control over the use of data", a reminder that the issues of autonomy and privacy are never far away.

Another option appears to be to move away from the tool paradigm by cultivating non-alienating human-machine synergy. For Villani (Villani, p. 18) this synergy could be based on developing innately human skills such as creativity, manual dexterity or problem solving. New ways to reach these types of goals are required: we need new means (Villani, p. 23) or digital literacy training from elementary school through university, for all citizens (CNIL, p. 54).

The CNIL (CNIL, p. 48) proposes a principle of loyalty which sums up the spirit of what sound autonomy management could look like in the age of AI. "A loyal algorithm should not incite, reproduce or reinforce any kind of discrimination whatsoever, even unknowingly, by designers". This loyalty must be understood as not only extending towards individual users, but society as a whole—because all of society could be affected by algorithmic "rulings" that are explicitly unwanted. We also see how issues of autonomy are often aligned with those of justice.

For its part, the preliminary version of the Montréal Declaration suggests the following principle: "AI development should foster the autonomy of all human beings and control the autonomy of computer systems." Because of its very general nature, this principle is in keeping with the other reports. It sets itself apart slightly by introducing the autonomy of computer systems—whereas other reports focus on human autonomy and the risks of it dwindling.

## JUSTICE

Every report contains recommendations on justice, with the main themes being social justice, human skills, human-AI synergy, AI skills and respecting humans.

The key idea is that artificial intelligence, and the systems that use its power, must lead to a fairer, more equitable society (AI Now, p. 2). This idea is rooted in two principles:

1. **The goal of AI must be to redress the shortcomings of society in these fields (UKRS, p. 12);**
2. **we must be careful, especially during the development and deployment phases, not to create or perpetuate injustice (Google). These two goals can be reached by providing solutions at many different levels.**

AI innovations must benefit everyone (Google). The idea is a trickle-down effect (Villani, p. 16): benefits (in service) and wealth (in knowledge, in technology/ technique, in accumulated data) must not be reserved for large private companies (Villani, p. 12) or the upper echelons of society—who may represent a majority of the population in terms of culture, religion or race, or a minority of the population in terms of income, such as the "1%". (Villani, p. 22).

AI innovations must aim for a better world where existing inequalities are addressed and fought in the legal system (Asilomar), in access to health care, or in protecting usually overlooked populations (AI Now, pp. 1 and 2; Villani, p. 18; Google). A national database that helps to objectively identify inequalities between men and women in the workplace (Villani, p. 23) needs to be created to resolve gender-based discrimination issues. Likewise, we must steer AI development toward applications that help improve both economic performance and the common good.

For everyone to benefit from AI, it must be inclusive, at every level (Villani, p. 19). This means that at every stage, from design to deployment to maintenance, an AI system should be examined by public authorities. Incentive policies are also needed to include underrepresented populations such as women or minorities.

Additional training in social sciences and ethics can help sensitize designers to these issues and provide the conceptual and intellectual tools to address them (AI Now, pp. 1 and 2). Likewise, research on algorithm interpretability and robustness as well as issues of equality, privacy and causality must be promoted and funded (UKRS, p. 13).

Lastly, justice also concerns legal institutions that can be directly affected by AI development. Here is what the different reports propose:

- > A legal framework must be developed to guarantee social justice, ensure everyone is represented when designing and using algorithms, reduce inequalities, and prevent abuse or misuse that could arise with unregulated AI use (Asilomar).
- > An important overhaul of the judicial system on all matters pertaining to artificial intelligence and data is overdue, especially for questions of sovereignty, ownership, data citizenship and governance (UKRS, p. 12; Asilomar; IEEE, p. 22). Likewise, we must give considerable thought to the notion of transparency and its evaluation criteria if we wish to assess the compliance of companies using AI systems (IEEE, p. 30).
- > These legal and ethical frameworks should be designed with the buy-in of all stakeholders in society: the scientific community, public authorities, industry players, entrepreneurs and civil society organizations (Villani, p. 21). Control systems should be regularly evaluated to ensure they are satisfactorily fulfilling their mission.
- > In the same way that it was determined that a company is a separate legal entity, we must reflect on the legal nature of AI itself (Asilomar).
- > When artificial intelligence is involved in legal decisions, auditing, interpretation, verification and explanation measures must be implemented (Asilomar).

For its part, the preliminary version of the Montréal Declaration suggests the following principle: "AI development must promote justice and seek to eliminate discrimination, namely that of gender, age, mental and physical abilities, sexual orientation, ethnic and social origins and religious beliefs."

This statement primarily addresses social justice and problems of equality and equity whether by redressing past discrimination or anticipating future discrimination. The Montréal Déclaration does not specify how to achieve these goals, unlike many reports that suggest, for example, more inclusion and social representation in the early phases of designing artificial intelligence systems. Moreover, it does not discuss implications that are specific to the legal system.

## PRIVACY

Explicit recommendations on privacy are contained in every report, with the exception of AI Now. These recommendations are namely associated with issues of privacy and confidentiality, collectivism/individualism, digital commons, governance, social justice, transparency, safety and system integrity.

On a very general level, the issue of privacy expresses the idea that the user should have control over their data—a link can therefore be made with autonomy. Asilomar, for example, maintains that people should have the right to access, manage and control the data they generate, while Google claims that protection of privacy should play an important role in the design of AI principles and AI system development. We note, however, that the reports provide few details as to general privacy principles. The issue appears difficult to address in such a general manner.

Protection of privacy implies various governance frameworks, namely regulatory and standard-setting bodies (IEEE, p. 22). For the CNIL (CNIL, p. 45), the law is responsible for overseeing the use of personal data by AI. A pertinent example is offered by (Villani, p. 11) who, in the wake of the General Data Protection Regulation (GDPR), mentions the right to data portability, meaning individuals' rights to recover the data they generate on one platform and use on another platform.

Two trends seem to emerge in the socio-political models that determine data governance. Villani and the CNIL seem to adopt the logic of data as a

common good, while the UKRS appears to align itself with a more “liberal” logic, or at least one more centred on the individual. Once again, we need to be careful in contrasting these approaches as it is difficult to infer a general trend from a few recommendations. Villani (Villani, p. 11) asks government authorities impose “openness on certain data of public interest”. We need only think of medical data that, when pooled together, could help advance research and benefit an entire population, or environmental data, for example, which could help collectively fight climate change. This suggestion echoes that of the CNIL (CNIL, p. 59), which suggests that the state launch a “major research program based on data contributed by citizens exercising their right to data portability among private stakeholders.”

For the UKRS, protecting privacy in scientific research takes precedence. The issue is of protecting individuals, which is why researchers should keep track of potential future uses for the data they collect, and integrate this aspect into the consent participants provide for research (UKRS, p. 8). This concern must be present from the moment data is collected until it is potentially shared or redistributed. The contrast between the two types of logic is not that significant, as the CNIL also suggests developing research infrastructure that is “respectful of personal data” (CNIL, p. 59), while the UKRS is not opposed to the logic of a “data commons” when it is generated by studies funded by public funds or charities (UKRS, p. 8).

We conclude by noting that there is an evident link between the issues of protection of privacy and justice since personal data could serve as the basis for discriminatory policies. This aspect is present in most reports.

For its part, the preliminary version of the Montréal Declaration proposes the following principle: “AI development should guarantee the respect of privacy and allow those who use it to access their personal data as well as the kinds of information used by the algorithm.” Although this principle provides a summary that reflects the findings of other reports, it does not provide an exhaustive overview of the complex and wide-ranging subject of privacy. In particular, this principle of the Montréal

Declaration does not broach issues of transparency which, in many ways, are the corollary of privacy—and are analyzed in the next section on knowledge.

## KNOWLEDGE

Recommendations on knowledge are in every report. The most common themes are social justice, transparency, human skills and digital literacy.

The two main areas of focus are increasing knowledge of the public and the authorities that will validate or verify AI systems. The autonomy, and transparency, of public and governing bodies can only exist if we let the public and the government exercise it by providing the necessary mechanisms and infrastructure as well as training, education and critical thinking.

A new digital literacy must be developed to hone critical thinking and understanding of these new technologies (CNIL, p. 54), from grade school through university, for all citizens. To achieve this, we must encourage new ways of thinking about intellectual autonomy and get people involved in thinking about everyday AI issues (CNIL, p. 57)—for example, understanding what it means when you give your consent. In other words, we must address the asymmetry that exists between AI service providers and users/citizens.

To protect the public, we need to think about possible ways AI systems could be used for harm. This implies establishing the framework for an educational method and appropriate measuring tools (IEEE, p. 31); for example, a validation test in schools. Again, notions of ethics and social sciences are suggested to complement this learning (IEEE, p. 31). People who are most “at risk”, meaning those identified as being especially gullible and/or who would suffer the worst consequences from harmful uses are those to be targeted as priorities (IEEE, p. 31).

A number of recommendations, in addition to those intended for the public, are addressed to government agents, elected representatives who vote on laws,

the legal system that will enforce them and the institutions that will serve as safeguards (IEEE, p. 31). Other "at risk" sectors, such as medicine, human resources (recruitment) or even marketing should be especially vigilant (CNIL, p. 55).

Obviously, algorithm and AI system designers are also affected by these measures: their training should be complemented by studies in humanities so that they grasp the social and economic issues of the solutions they come up with and understand the impact their solutions may have in practice (CNIL, p. 55). Many reports recommend reinforcing cultural, social and gender diversity; the idea is that by multiplying representatives of society at every level of AI design, we will have a better understanding of all the parameters, context and viewpoints to take into consideration (CNIL, p. 55).

Numerous recommendations underscore the knowledge required to correctly operate the AI control systems and evaluation infrastructure. We first need to establish standards and regulatory bodies to monitor the various steps of the design process for AI systems and ensure they respect human rights, freedoms, dignity, privacy and traceability (IEEE, p. 22). These standards must be implemented by public institutions (IEEE, p. 30) that develop transparent measuring tools that are accessible to the public (AI Now, p. 1) and designed by impartial experts and professionals.

One recurrent recommendation in these reports is the need for transparent regulatory bodies. Giving the public access to all these evaluation methods will allow them not only exercise, but demonstrate their knowledge. With trained and motivated users and systems led by transparent and authoritative committees, the last step would be to leave citizens free to experiment, deploy their digital literacy and exercise their critical thinking. It suggests, for example, that the various user platforms for AI systems offer information on how their algorithms operate (CNIL, pp. 45 and 48). Specific information on the data used and algorithm logic could be made available on user profile pages (CNIL, p. 56). To promote understanding, users should be able to "play" with the system by changing parameters (CNIL, p. 57).

One last point: we must ensure the transition by verifying and improving training in schools. Digital literacy is defined in different ways, from the aforementioned ethics and critical thinking to the knowledge of key principles of programming or machine learning (UKRS, p. 9). Once again, it is a matter of addressing the information asymmetry that can exist between users, developers and citizens. Governments, experts in mathematics and programming, companies and education professionals must all contribute to this digital literacy in order to build a much-needed and sufficient knowledge base (UKRS, p. 9). Many recommendations highlight the importance and interest of addressing notions of ethics, social sciences and public health in educational activities (UKRS, p. 9).

As for the educational system, its mission should also be to train a new generation of workers and researchers with the skills required to navigate a world of AI systems. Not only should we reconsider the initial training offered in university, but we must also provide ongoing training and new skill sets to workers whose tasks will be drastically altered. These recommendations are all the more meaningful in the context of job insecurity caused by machines replacing humans (UKRS, p. 9). Both universities and industries must reflect on future needs in terms of skills, from machine learning to the science of data (UKRS, p. 9).

On the subject of knowledge, the preliminary version of the Montréal Declaration proposes the following principle: "AI development should promote critical thinking and protect us from propaganda and manipulation." If awakening critical thought echoes the notions of digital literacy found to varying degrees and ways in the reports, the Montréal Declaration focuses on protecting the public from propaganda and manipulation, whereas the notions of accomplishment, freedom, and power are more evident in other reports. For many, knowledge not only offers protection, but opens the door to many possibilities.

## DEMOCRACY

The value (or notion) of democracy is apparent in all reports, in comparable proportions. The recommendations that address democracy are associated with governance, collectivism/individualism, democratic governance, digital commons, privacy and confidentiality, in particular.

The first theme deals with governance. As we have already seen with the principle of autonomy, the reports insist that AI remain under human control (AI Now, p. 1), which explains the need for a specialized supervision framework (IEEE, p. 22; UKRS, p. 12) or audit systems (CNIL, p. 57). Can we allow the private sector to self-regulate? The answer that emerges from these reports is rather pessimistic—but we must admit that the opposing viewpoint is essentially nonexistent, as the only company whose principles/recommendations are available (Google) does not address the question. As for the type of governance, certain recommendations use a relatively classic top-down logic: IEEE or UKRS, for example, with the idea of seeking social acceptability or “consulting” citizens (IEEE, p. 31). Asilomar discusses the need for dialogue between researchers and policy-makers. The more radical or direct conceptions of democracy do not appear explicitly in the reports.

Regardless, everyone agrees that AI development must be regulated—Villani even specifies that, for example, a special framework must be developed to protect the most sensitive sets of data (Villani, p. 20). But what lends these recommendations a truly democratic dimension is that the framework or control in question must be transparent. AI Now advocates that AI systems used by public agencies be subject to public audits, tests and revisions (AI Now, p. 1). The idea of a “public body of experts” that would control the algorithms “to verify for example that “they do not discriminate” is also echoed by the CNIL (CNIL, p. 58), which goes even further than AI Now because the mission of these experts does not appear limited to the public sector. The issue of algorithmic opacity is often mentioned, since it hinders transparency. Villani notes that being able to “open the black boxes” is a democratic issue (Villani, p. 21). Support for research in the field of algorithm explainability, therefore, is necessary (Villani, p. 21).

Democracy is also present in calls for diversity—cultural, social and gender, as specified by CNIL (CNIL, p. 55)—among algorithm developers since it is unlikely that a sub-group (usually of rich white men) can adequately anticipate and respond to the needs of all members of society. Villani, therefore, promotes “inclusive and diverse” AI (Villani, p. 22) whereas the IEEE (IEEE, p. 27) recommends that designers and developers be aware of the diverse cultural norms that exist among AI system users. Finally, for Google, one of the roles of companies is to share knowledge and thereby democratize AI so that more people can develop useful applications for it (Google).

The preliminary version of the Montréal Declaration proposes the following principle: “AI development should foster informed participation in public life, cooperation and democratic debate.” The lack of issues on diversity is unsurprising because they are addressed in the principle of justice in the Montréal Declaration. It is worth considering, however, whether issues of governance and transparency should accompany this principle. The principle of democracy in the Montréal Declaration says nothing about who should control AI development and how sharing between public and private governance, experts and laypeople should be addressed.

## RESPONSIBILITY

All reports contained recommendations on responsibility. The most predominant themes were safety and system integrity, social justice, AI skills, sharing responsibility, accountability and shared responsibility.

The issue of decision-making first touches upon the notion of responsibility: when AI can act alone, when should it be supervised or completed by a human being (AI Now, p. 1)? For some, a machine must never make a decision by itself (meaning with no human intervention) if there are serious consequences for people (CNIL, p. 45).

To correctly assign responsibility to one entity or the other (or both), we must ensure that the humans who interact with AI have the necessary



training to understand, think critically and measure the limits and biases that need to be corrected. Some recommendations go one step further by suggesting that the moment AI shows biases and discrimination, and as its presence in our social and economic lives increases, “opening the black box” becomes a democratic issue (Villani, p. 18). Because competitiveness issues foresee companies not all nor always being completely transparent, it is often recommended that the use of AI in the public sphere be as transparent as possible. First, by not relying on private companies to manage public systems (AI Now, p. 1), then by subjecting public systems to the strictest tests, evaluations, audits, inspections, and responsibility standards (AI Now, p. 1).

Being responsible also means anticipating problems: how can we avoid hurdles, what infrastructures should we implement? Some recommendations on this issue are clear: the ruling principle should be vigilance (CNIL, p. 50) and AI designers should always remember that algorithms can be unpredictable and they are autonomous and constantly evolving. This principle of vigilance seeks to temper, or at least counterbalance the risk of excessive trust in AI (CNIL, p. 50). Many potential solutions are proposed, such as devising recording and traceability systems to go back to the source of an algorithm and determine responsibility in case of a problem (IEEE, p. 27).

Each report highlights the fact that currently, the legal system can barely keep up with the frenzied pace of data and AI development, and consequently, with ways to regulate these new technologies. Resources need to be mobilized so that it can stay apace (Asilomar).

Two key elements appear necessary to guide AI systems:

1. **Involving the judicial system to control, correct, delineate and help;**
2. **involving independent scientists in the design of devices to monitor, call and label all these AI systems. Both groups will have to work together to establish good control test practices (Asilomar).**

Companies must not remain passive, however. Since they must ensure not to amplify biases or make mistakes (AI Now, p. 1), a significant share of the work to be done is in prevention, namely by using test versions before the global launch of an AI app (AI Now, p. 1). These preliminary tests must not only verify the way the algorithms are built but, above all, verify the data on which they were trained (AI Now, p. 1). This is why it is recommended to have information on the source and management of this training data, as well as back-ups so they can be explored in case of an anomaly (AI Now, p. 1).

Responsibility in the legal field is a hot topic, and the responsibility for making the most appropriate decision and preventing injustice (creating it, reinforcing it) is at the heart of many discussions. Hence, Asilomar recommends that any autonomous system involved in legal decisions be able to provide clear explanations on its decision-making process (Asilomar). These explanations should be analyzed by a competent person with adequate training to understand the workings of the algorithm and offer an intelligible explanation.

The theme of responsibility also concerns mediation between the public and AI system providers, therefore openness and transparency. We must include all members of society in the debate on human responsibility (Villani, p. 22). The public will be asked to think critically in mediation cases (previously discussed)—if the public wishes to defend itself in case of a dispute, the algorithms must be explainable, and the public must be able to understand them—, as well as in public and citizen consultations or open national audits.

For its part, the preliminary version of the Montréal Declaration proposes the following principle: “The different stakeholders in AI development should assume responsibility by minimizing the risks of these technological innovations.” The Montréal Declaration sums up the essence of the recommendations made in the various reports, but remains very general (as the reports can offer more detailed recommendations). The Montréal Declaration, therefore, could help clarify the role of the many different stakeholders involved in building these systems, the impact of their work on each other and the pitfalls that should be avoided.

## 3. REPORTS ON AI DEVELOPMENT: TECHNICAL DATA SHEETS

### 3.1 THE SEVEN REPORTS STUDIED

#### (AI NOW) AI NOW 2017 REPORT

Subtitle: no

Published: November 2017

Country: USA

Language: English

Organization or signatories: AI Now Institute (report signed by Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker, Kate Crawford)

Number of pages: 37

Summary: yes (3 pages)

Well identified general ethical principles: no

Well identified recommendations: yes (10)

Main themes: work and automation, biases and inclusion, rights and freedoms, ethics and governance.

Notes: An annual report that quotes many recent studies and is devoted to updating people on advances in research.

Link: [https://ainowinstitute.org/AI\\_Now\\_2017\\_Report.pdf](https://ainowinstitute.org/AI_Now_2017_Report.pdf)

#### (CNIL) HOW CAN HUMANS KEEP THE UPPER HAND? THE ETHICAL MATTERS RAISED BY ALGORITHMS AND ARTIFICIAL INTELLIGENCE

Subtitle: Report on the public debate led by the French data protection authority (CNIL) as part of the ethical discussion assignment set by the Digital Republic bill

Published: December 2017

Country: 80

Language: English (translated from French :

Comment permettre à l'homme de garder la main -

Les enjeux éthiques des algorithmes et de l'IA)

Organization or signatories: CNIL: Commission nationale informatique et liberté (foreword by Isabelle Falque-Pierrotin, president of the CNIL)

Number of pages: 80

Summary: yes (2 pages)

Well identified general ethical principles: yes (vigilance and loyalty)

Well identified recommendations: yes (6)

Main themes: ethical uses of AI, applications for each field (health care, education, living in society and politics, culture and media, justice, banks and finance, safety and defence, insurance, work and HR).

Notes: One of the most thorough reports on the ethical issues of AI.

Link: [https://www.cnil.fr/sites/default/files/atoms/files/cnil\\_rapport\\_garder\\_la\\_main\\_web.pdf](https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf)

#### (IEEE) ETHICALLY ALIGNED DESIGN. VERSION 2—FOR PUBLIC DISCUSSION

Subtitle: A vision for prioritizing human well-being with autonomous and intelligent systems.

Published: December 2017

Country: international

Language: English

Organization or signatories: IEEE (Institute of Electrical and Electronics Engineers); signed by IEEE subcommittees that regroup several hundred international participants.

Number of pages: 266

Summary: yes (17 pages)

Well identified general ethical principles: yes (5)

Well identified recommendations: yes

Main themes: ethical, legal, political issues; questions specifically tied to information and communication technologies; safety; ethics by design; data control.

Notes: Each chapter was written by committees of experts.

Link: <https://ethicsinaction.ieee.org/>



## (ASILOMAR) ASILOMAR AI PRINCIPLES

Subtitle: no

Published: 2017

Country: international

Language: English with Chinese; German, Japanese, Korean and Russian translations available

Organization or signatories: Future of Life Institute, signed by over 1200 researchers and 2500 non-researchers.

Number of pages: 2

Summary: no

Well identified general ethical principles: yes (23)

Well identified recommendations: no

Main themes: ethics of research, moral values, long-term issues.

Notes: It is not a report, but a series of principles that stem from discussions between experts during a conference in Asilomar, California. In 1975, another conference in Asilomar established bioethics principles.

Link: <https://futureoflife.org/ai-principles/?cn-reloaded=1>

## (UKRS) AI IN THE UK: READY, WILLING, AND ABLE?

Subtitle: no

Published: April 16, 2018

Country: UK

Language: English

Organization or signatories: Parliament (House of Lords); 13-person committee.

Number of pages: 184

Summary: yes (5 pages)

Well identified general ethical principles: no

Well identified recommendations: yes (73)

Main themes: questions of ethics and economics policy ("innovation in AI"). Impact of AI on different fields: economy, work, education, health care, justice.

Notes: The report is divided into 420 paragraphs, with the author usually identified in the notes.

Link: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>

## (VILLANI) DONNER UN SENS À L'INTELLIGENCE ARTIFICIELLE

Subtitle: Pour une stratégie nationale et européenne

Published: March 8, 2018

Country: France

Language: French

Organization or signatories: Parliamentary missions entrusted to MP Cédric Villani and six (6) other members of parliament.

Number of pages: 235

Summary: yes (15 pages)

Well identified general ethical principles: no

Well identified recommendations: no

Main themes: questions of ethics and political economy, research policies, impact on work and education sectors, health, agriculture, transportation and defence.

Link: <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/184000159.pdf>

## (GOOGLE) AI AT GOOGLE: OUR PRINCIPLES

Subtitle: no

Published: June 7, 2018

Country: USA

Language: English

Organization or signatories: Google, presented by its CEO Sundar Pichai

Number of pages: 3

Summary: no

Well identified general ethical principles: yes (7)

Well identified recommendations: yes (4)

Main themes: AI ethics

Notes: The company commits to not deploying AI in certain fields (weapons) or circumstances (against human rights).

Link: <https://www.blog.google/technology/ai/ai-principles/>

## 3.2

### REPORTS EXAMINED, BUT NOT SELECTED

#### A NEXT GENERATION ARTIFICIAL INTELLIGENCE DEVELOPMENT PLAN

Subtitle: no

Published: July 2017

Country: China

Language: English (translation)

Number of pages: 28

Summary: no

Well identified general ethical principles: no

Well identified recommendations: yes

Main themes: national strategy for economic development

Link: <https://chinacopyrightandmedia.wordpress.com/2017/07/20/a-next-generation-artificial-intelligence-development-plan/>

#### STRATEGY FOR DENMARK'S DIGITAL GROWTH

Subtitle: no

Published: 2018

Country: Denmark

Language: English

Organization or signatories: Ministry of Industry, Business and Financial Affairs

Number of pages: 68

Summary: yes (6 pages)

Well identified general ethical principles: no

Well identified recommendations: yes

Main themes: national strategy for economic development

Link: <https://em.dk/english/news/2018/01-30-new-strategy-to-make-denmark-the-new-digital-frontrunner>

#### COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS

Subtitle: Artificial Intelligence for Europe

Published: April 25, 2018

Country: European Union

Language: English

Organization or signatories: European Commission

Number of pages: 20

Summary: no

Well identified general ethical principles: yes

Well identified recommendations: yes

Main themes: national strategy for economic development

Link: <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>

#### FINLAND'S AGE OF ARTIFICIAL INTELLIGENCE

Subtitle: Turning Finland into a leading country in the application of artificial intelligence: Objective and recommendations for measures

Published: December 18, 2017

Country: Finland

Language: English

Organization or signatories: Ministry of Economic Affairs and Employment

Number of pages: 76

Summary: yes (3 pages)

Well identified general ethical principles: no

Well identified recommendations: yes (8)

Main themes: national strategy for economic development

Link: [http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap\\_47\\_2017\\_verkkojulkaisu.pdf?sequence=1&isAllowed=y](http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y)

## ETHICS COMMISSION AUTOMATED AND CONNECTED DRIVING

Subtitle: no

Published: June 2017

Country: Germany

Language: English

Organization or signatories: Federal Ministry of Transport and Digital Infrastructure

Number of pages: 36

Summary: no

Well identified general ethical principles: no

Well identified recommendations: yes

Main themes: Ethics of self-driving vehicles

Link: [https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?\\_\\_blob=publicationFile](https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile)

## NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE #AIFORALL

Subtitle: Discussion paper

Published: June 2018

Country: India

Language: English

Organization or signatories: NITI Aayog

Number of pages: 115

Summary: yes (3)

Well identified general ethical principles: yes

Well identified recommendations: yes

Main themes: national strategy for economic and societal development

Link: [http://niti.gov.in/writereaddata/files/document\\_publication/NationalStrategy-for-AI-Discussion-Paper.pdf](http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf)

## ARTIFICIAL INTELLIGENCE AT THE SERVICE OF CITIZENS

Subtitle: no

Published: March 2018

Country: Italy

Language: English

Organization or signatories: The Agency for Digital Italy

Number of pages: 79

Summary: yes (5 pages)

Well identified general ethical principles: yes

Well identified recommendations: yes

Main themes: AI's impact on society and the public administration to promote change

Link: <https://ia.italia.it/en/assets/whitepaper.pdf>

## ARTIFICIAL INTELLIGENCE TECHNOLOGY STRATEGY

Subtitle: Report of Strategic Council for AI Technology

Published: March 31, 2017

Country: Japan

Language: English

Organization or signatories: Strategic Council for AI Technology

Number of pages: 25

Summary: no

Well identified general ethical principles: no

Well identified recommendations: yes

Main themes: national strategy for economic development

Link: <http://www.nedo.go.jp/content/100865202.pdf>

## TOWARDS AN AI STRATEGY IN MEXICO

Subtitle: Harnessing the AI Revolution

Published: June 2018

Country: Mexico

Language: English

Organization or signatories: British Embassy in Mexico through the Prosperity Fund, Oxford Insights, C Minds

Number of pages: 52

Summary: yes (3 pages)

Well identified general ethical principles: no

Well identified recommendations: yes (21)

Main themes: national strategy for economic development

Link: [https://docs.wixstatic.com/ugd/7be025\\_e726c582191c49d2b8b6517a590151f6.pdf](https://docs.wixstatic.com/ugd/7be025_e726c582191c49d2b8b6517a590151f6.pdf)

## SHAPING A FUTURE NEW ZEALAND

Subtitle: An Analysis of the Potential Impact and Opportunity of Artificial Intelligence on New Zealand's Society and Economy

Published: May 2018

Country: New Zealand

Language: English

Organization or signatories: AI Forum of New Zealand

Number of pages: 108

Summary: yes (5 pages)

Well identified general ethical principles: yes

Well identified recommendations: yes (14)

Main themes: national strategy for economic development

Link: <http://resources.aiforum.org.nz/AI+Shaping+A+Future+New+Zealand+Report+2018.pdf>

## ARTIFICIAL INTELLIGENCE IN SWEDISH BUSINESS AND SOCIETY

Subtitle: Analysis of development and potential

Published: May 2018

Country: Sweden

Language: English

Organization or signatories: Vinnova

Number of pages: 32

Summary: no

Well identified general ethical principles: no

Well identified recommendations: yes

Main themes: economic development and public services

Link: [https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr\\_18\\_09.pdf](https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr_18_09.pdf)

## INDUSTRIAL STRATEGY

Subtitle: AI Sector Deal

Published: April 2018

Country: UK

Language: English

Organization or signatories: Government

Number of pages: 21

Summary: yes (3 pages)

Well identified general ethical principles: no

Well identified recommendations: yes

Main themes: national strategy for economic

development

Link: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/702810/180425\\_BEIS\\_AI\\_Sector\\_Deal\\_\\_4\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal__4_.pdf)

## PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE

Subtitle: no

Published: October 2016

Country: USA

Language: English

Organization or signatories: Executive Office of the President, National Science and Technology Council Committee on Technology

Number of pages: 58

Summary: yes (4)

Well identified general ethical principles: yes

Well identified recommendations: yes (23)

Main themes: current state of AI, present and future applications, questions raised for society

Link: [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf)

## THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

Subtitle: no

Published: October 2016

Country: USA

Language: English

Organization or signatories: National Science and Technology Council, Networking and Information Technology Research and Development Subcommittee

Number of pages: 48

Summary: yes (2 pages)

Well identified general ethical principles: no (a few)

Well identified recommendations: yes (7)

Main themes: objectives for AI research funded by the federal government

Link: [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/national\\_ai\\_rd\\_strategic\\_plan.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf)

## ARTIFICIAL INTELLIGENCE, AUTOMATION, AND THE ECONOMY

Subtitle: no

Published: December 2016

Country: USA

Language: English

Organization or signatories: Executive Office of the President

Number of pages: 55

Summary: yes (4 pages)

Well identified general ethical principles: no

Well identified recommendations: yes (3)

Main themes: impact of AI automation on the economy and strategies to increase the benefits

Link: <https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>

## SUMMARY OF THE 2018 WHITE HOUSE SUMMIT ON ARTIFICIAL INTELLIGENCE FOR AMERICAN INDUSTRY

Subtitle: no

Published: May 10, 2018

Country: USA

Language: English

Organization or signatories: The White House Office of Science and Technology Policy

Number of pages: 15

Summary: yes (1 page)

Well identified general ethical principles: no

Well identified recommendations: yes

Main themes: national strategy for economic development

Link: <https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf>

## 3.3

### OTHER REPORTS CONSULTED

(Sweden) National Approach for Artificial Intelligence  
[https://www.regeringen.se/49a828/contentassets/844d30fb0d594d1b9d96e2f5d57ed14b/2018ai\\_webb.pdf](https://www.regeringen.se/49a828/contentassets/844d30fb0d594d1b9d96e2f5d57ed14b/2018ai_webb.pdf)

(Germany) Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz  
[https://www.bmwi.de/Redaktion/DE/Downloads/E/eckpunktepapier-ki.pdf?\\_\\_blob=publicationFile&v=4](https://www.bmwi.de/Redaktion/DE/Downloads/E/eckpunktepapier-ki.pdf?__blob=publicationFile&v=4)

(Finland) Work in the Age of Artificial Intelligence  
[http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160931/19\\_18\\_TEM\\_Tekoalyajan\\_tyo\\_WEB.pdf](http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160931/19_18_TEM_Tekoalyajan_tyo_WEB.pdf)

(China) Three-Year Action Plan to Promote the Development of New-Generation Artificial Intelligence Industry  
<http://www.miit.gov.cn/n1146295/n1652858/n1652930/n3757016/c5960820/content.html>

(Australia) Australia 2030: Prosperity Through Innovation  
<https://www.industry.gov.au/sites/g/files/net3906/f/May%202018/document/pdf/australia-2030-prosperity-through-innovation-full-report.pdf>

(Thanks to Paloma Fernandez-McAuley for her help.)

# 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



CENTRE DE RECHERCHE EN ETHIQUE



**CIFAR**  
AI &  
Society  
Program



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



**Canada** 



Centre  
Culturel  
Canadien  
Paris

Canadian  
Cultural  
Centre  
Paris







</>

[montrealdeclaration-responsibleai.com](https://montrealdeclaration-responsibleai.com)